

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2002, 10:12:11 ; Search time 30.32 Seconds  
(without alignments)  
1161.293 Million cell updates/sec

Title: US-09-903-925a-263  
Perfect score: 1720  
Sequence: 1 MVSGAPALGGGCLGFTS.....AQQGALRAPSQSGGAARS 317

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: A\_Geneseq\_032802:\*  
1: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1720	100.0	317	20 AAY06482	Human tumour-assoc
2	1720	100.0	317	20 AAY13391	Amino acid sequenc
3	1720	100.0	317	21 AAB11700	Human serine prote
4	1720	100.0	317	21 AAY93689	Amino acid sequenc
5	1720	100.0	317	22 AAY72890	Human serine prote
6	1720	100.0	317	22 AAB80259	Human PRO343 prote
7	1720	100.0	325	22 AAU23215	Novel human enzyme
8	1720	100.0	325	22 AAU17037	Human novel secret
9	1720	100.0	351	22 AAU16966	Human novel secret
10	1714	99.7	322	22 AAU23751	Novel human enzyme
11	1714	99.7	322	22 AAU17038	Human novel secret

12	1706	99.2	319	21 AAB11701	Human serine prote
13	1626	94.5	306	21 AAB11702	Human serine prote
14	1484	86.3	271	19 AAW77302	Amino acid sequenc
15	1444.5	84.0	327	22 AAY72891	PFEX-C-E-HIS fusio
16	1310	76.2	302	21 AAB11708	Human serine prote
17	1157.5	67.3	308	21 AAB11709	Mouse serine prote
18	1074	62.4	234	21 AAB11706	Human serine prote
19	1047	60.9	207	21 AAB11704	Human serine prote
20	1017	59.1	231	22 AAG75586	Human colon cancer
21	739	43.0	290	21 AAY73388	HTRM clone 3376404
22	739	43.0	290	22 AAM39388	Human polypeptide
23	739	43.0	290	22 AAU12282	Human PRO4327 poly
24	739	43.0	290	22 AAE03821	Human gene 4 encod
25	739	43.0	290	22 AAB73945	Human protease T.
26	734	42.7	328	22 AAM41174	Human polypeptide
27	730	42.4	297	19 AAW77304	Amino acid sequenc
28	680.5	39.6	315	22 AAB73946	Fusion protein of
29	645	37.5	131	21 AAB11705	Human serine prote
30	628.5	36.5	284	22 AAU00467	Human serine prote
31	585.5	34.0	316	22 AAU00468	Recombinant human
32	582.5	33.9	319	21 AAB36481	Fusion gene with h
33	582.5	33.9	319	22 AAB67541	Amino acid sequenc
34	581	33.8	328	21 AAB36480	Fusion gene with h
35	581	33.8	328	22 AAB67540	Amino acid sequenc
36	554	32.2	276	19 AAW64242	Murine mast cell p
37	554	32.2	276	19 AAW63172	Mouse mast cell pr
38	549	31.9	658	22 AAE06934	Human membrane-typ
39	549	31.9	802	22 AAE06933	Human membrane-typ
40	547	31.8	802	20 AAY41710	Human PRO618 prote
41	547	31.8	802	21 AAB44266	Human PRO618 (UN03
42	547	31.8	802	21 AAB24052	Human PRO618 prote
43	538.5	31.3	274	19 AAW64234	Rat homologue of m
44	536.5	31.2	314	19 AAW77297	Amino acid sequenc
45	534	31.0	235	22 AAE06932	Human membrane-typ

ALIGNMENTS

RESULT 1	
AAY06482	
ID AAY06482 standard; Protein; 317 AA.	
XX AC AAY06482;	
XX DT 27-SEP-1999 (first entry)	
XX DE Human tumour-associated protein PRO343.	
XX KW PRO343; UNQ302; cancer; tumour; diagnosis; therapy; human.	
XX OS Homo sapiens.	
XX FH Key	Location/Qualifiers
FT Peptide	1..32 "signal peptide"
FT Protein	/note= "mature protein"
FT Modified-site	70
FT FT	/note= "N-glycosylated"
XX PN WO9935170-A2.	
XX PD 15-JUL-1999.	
XX PF 05-JAN-1999;	99WO-US00106.
XX PR 20-NOV-1998;	98US-0109304.
XX PR 05-JAN-1998;	98US-0070440.
XX PR 29-APR-1998;	98US-0083500.
XX PR 22-MAY-1998;	98US-0086414.
XX PR 10-JUN-1998;	98US-0088742.
XX PR 10-NOV-1998;	98US-0107783.

XX (GETH ) GENENTECH INC.  
 PA Botstein D, Goddard A, Gurney AL, Hillan KJ, Lawrence DA;  
 PI Roy MA, Wood WI;  
 XX WPI; 1999-430385/36.  
 DR N-PSDB; AAX87259.  
 XX  
 PT Antibody against proteins expressed in neoplastic cells, useful for  
 PT tumor diagnosis and treatment  
 XX  
 PS Example 1; Fig 12; 162pp; English.  
 XX  
 CC This sequence represents human PRO343 (UNQ302), a protein encoded  
 CC by the novel cDNA clone DNA43318 (see AAX87259). Amplification of  
 CC DNA43318 was observed in primary lung tumours and in primary colon  
 CC tumours, suggesting a significant role in tumour formation and  
 CC growth. Antagonists (e.g. antibodies) directed to PRO343 may have  
 CC use in cancer therapy. The invention identifies 14 genes (see  
 CC AAX87254-67) that are amplified in the genome of tumour cells. Such  
 CC amplification is expected to be associated with overexpression of  
 CC the gene product and to contribute to tumorigenesis. The encoded  
 CC proteins (see AAY06477-90) may be useful targets for the diagnosis  
 CC and/or treatment (including prevention) of certain cancers, and may  
 CC act as predictors of the prognosis of tumour treatment. Antibodies  
 CC that bind the proteins are claimed and used in claimed cancer  
 CC diagnostic kits.  
 XX  
 SQ Sequence 317 AA;  
 Query Match 100.0%; Score 1720; DB 20; Length 317;  
 Best Local Similarity 100.0%; Pred. No. 3.2e-147;  
 Matches 317; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MVVSGAPPALGGCGTFTSLLLASTAILNAARIPVPPACGKPOQLNRVVGEDSTDSE 60  
 DB 1 mvvsgappalggcgftftslilastailnaaripvpacgkppqqlnrvvvggedstdse 60  
 QY 61 WPVIVSIQKNGTHRCAGSLTSTRVITAACHFCNDLNKPYLFLSGLGAWOLGNPGRSQK 120  
 DB 61 wpviviqkngthrcagslltscrwvitaachfcdlnkpylflsvlgawqlngnpgsrsk 120  
 QY 121 VGVAVPEHPVYSKEGACADIALVRLESTQFSERVLPTCLPDASHLPPNTHCWISGW 180  
 DB 121 vgvavpehpvyskvegacadiavlrlestrqfservlplclpdasihlppnthcwisgw 180  
 QY 181 GSIDGVPPLPHTLQKLVPIIDSEVCSHLYWRGAGGPTEDMLCAGYLEGERDACLG 240  
 DB 181 gsidgvpplphpqclqkvlpiidsevcshlywrgaggptedmlcagylegerdaclg 240  
 QY 241 DSGGPLMCQVDGAWLLAGIISWGSGCAERNRPGVYIISLAHRSWKEIVQGVQLRGAQG 300  
 DB 241 dsggplmcqvdgawllagiiswsgcaernrpgvyislahrswekivqgvqlrgragg 300  
 QY 301 GGALRAPSQSGGAARS 317  
 DB 301 ggalrapsgsggaars 317  
 RESULT 2  
 AAY13391  
 ID AAY13391 standard; Protein; 317 AA.  
 XX  
 AC AAY13391;  
 XX  
 AC  
 XX  
 DT 25-JUN-1999 (first entry)  
 XX  
 DE Amino acid sequence of protein PRO343.  
 XX  
 KW Secreted protein; transmembrane protein; human; enterocolitis;  
 Zöllinger-Ellison syndrome; gastrointestinal ulceration;

KW congenital microvillus atrophy; skin disease; cell growth;  
 KW abnormal keratinocyte differentiation; psoriasis; epithelial cancer;  
 KW Parkinson's disease; Alzheimer's disease; ALS; neuropathy;  
 KW fibromodulin; dermal scarring; Usher Syndrome; Atrophia areata;  
 KW anti-thrombotic; wound healing; tissue repair.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO9914328-A2.  
 XX  
 PD 25-MAR-1999.  
 XX  
 PF 16-SEP-1998; 98WO-US19330.  
 XX  
 XX 25-NOV-1997; 97US-0066840.  
 PR 17-SEP-1997; 97US-0059113.  
 PR 17-SEP-1997; 97US-0059113.  
 PR 17-SEP-1997; 97US-0059117.  
 PR 17-SEP-1997; 97US-0059119.  
 PR 17-SEP-1997; 97US-0059121.  
 PR 17-SEP-1997; 97US-0059122.  
 PR 17-SEP-1997; 97US-0059184.  
 PR 18-SEP-1997; 97US-0059283.  
 PR 18-SEP-1997; 97US-0059286.  
 PR 15-OCT-1997; 97US-0062125.  
 PR 17-OCT-1997; 97US-0062285.  
 PR 17-OCT-1997; 97US-0062287.  
 PR 21-OCT-1997; 97US-0063486.  
 PR 24-OCT-1997; 97US-0062814.  
 PR 24-OCT-1997; 97US-0062816.  
 PR 24-OCT-1997; 97US-0063045.  
 PR 24-OCT-1997; 97US-0063120.  
 PR 24-OCT-1997; 97US-0063121.  
 PR 24-OCT-1997; 97US-0063127.  
 PR 27-OCT-1997; 97US-0063128.  
 PR 27-OCT-1997; 97US-0063329.  
 PR 27-OCT-1997; 97US-0063327.  
 PR 28-OCT-1997; 97US-0063541.  
 PR 28-OCT-1997; 97US-0063542.  
 PR 28-OCT-1997; 97US-0063544.  
 PR 28-OCT-1997; 97US-0063549.  
 PR 28-OCT-1997; 97US-0063550.  
 PR 28-OCT-1997; 97US-0063564.  
 PR 29-OCT-1997; 97US-0063435.  
 PR 29-OCT-1997; 97US-0063704.  
 PR 29-OCT-1997; 97US-0063732.  
 PR 29-OCT-1997; 97US-0063738.  
 PR 29-OCT-1997; 97US-0063734.  
 PR 29-OCT-1997; 97US-0064215.  
 PR 31-OCT-1997; 97US-0063735.  
 PR 31-OCT-1997; 97US-0063870.  
 PR 31-OCT-1997; 97US-0064103.  
 PR 03-NOV-1997; 97US-0064248.  
 PR 07-NOV-1997; 97US-0064809.  
 PR 12-NOV-1997; 97US-0065186.  
 PR 17-NOV-1997; 97US-0065846.  
 PR 18-NOV-1997; 97US-0065693.  
 PR 21-NOV-1997; 97US-0066120.  
 PR 21-NOV-1997; 97US-0066364.  
 PR 24-NOV-1997; 97US-0066772.  
 PR 24-NOV-1997; 97US-0066466.  
 PR 24-NOV-1997; 97US-0066770.  
 PR 24-NOV-1997; 97US-0066511.  
 PR 24-NOV-1997; 97US-0066453.  
 XX  
 PA (GETH ) GENENTECH INC.  
 XX  
 XX Chen J, Goddard A, Gurney AL, Pennica D, Wood WI, Yuan J;  
 PI WPI; 1999-229533/19.  
 XX N-PSDB; AAX52262.  
 DR  
 DR  
 XX New isolated human genes and polypeptides used in, e.g. treatment of



```

RESULT      4
AA93689
ID  AA93689 standard; Protein; 317 AA.
XX
XX
AC  AA93689;
XX
XX  03-OCT-2000 (first entry)
XX
XX  Amino acid sequence of novel polypeptide PRO343.
XX
XX  PRO201; PRO327; PRO1265; PRO344; PRO343; PRO347; PRO357;
XX  PRO715; PRO1017; PRO1112; PRO509; PRO853; PRO882; tumour cell;
XX  tumourigenesis; cancer; neoplastic cell growth; cell proliferation.
XX
XX  Homo sapiens.
XX
XX  Key      Location/Qualifiers
XX  Peptide  1..32
XX  Region   /note= "signal sequence"
XX  Modified-site 5..11
XX  Modified-site /note= "N-myristoylation site"
XX  Modified-site /note= "N-myristoylation site"
XX  Modified-site /note= "N-myristoylation site"
XX  Modified-site /note= "N-myristoylation site"
XX  Modified-site /note= "N-myristoylation site"
XX  Modified-site /note= "N-myristoylation site"
XX  Modified-site /note= "N-glycosylation site"
XX  Modified-site /note= "N-myristoylation site"
XX  Active-site /note= "active site for serine protease"
XX  Modified-site /note= "N-myristoylation site"
XX  Region     /note= "N-myristoylation site"
XX  Modified-site /note= "glycosylation attachment site"
XX  Modified-site /note= "N-myristoylation site"
XX  Modified-site /note= "N-myristoylation site"
XX
XX  WO200037640-A2.
XX
XX  29-JUN-2000.
XX
XX  16-DEC-1999; 99WO-US30095.
XX
XX  22-DEC-1998; 98US-0113296.
XX  08-MAR-1999; 99WO-US05028.
XX  02-JUN-1999; 99WO-US12252.
XX  01-SEP-1999; 99WO-US20111.
XX  15-SEP-1999; 99WO-US21090.
XX  30-NOV-1999; 99WO-US28313.
XX  01-DEC-1999; 99WO-US28409.
XX  02-DEC-1999; 99WO-US28301.
XX
XX  (GETH ) GENENTECH INC.
XX
XX  Botstein D, Goddard A, Gurney AL, Hillan K, Lawrence DA, Roy MA;
XX  Wood WI;
XX
XX  WPI: 2000-452188/39.
XX  N-PSDB; AAA46914.

```

```

XX  New anti-polypeptide antibody useful in the treatment and diagnosis of
XX  neoplastic cell growth and proliferation -
XX
XX  Claim 61; Fig 12; 220pp; English.
XX
XX  The present sequence represents a novel human polypeptide. The
XX  specification describes novel polypeptides designated PRO201, PRO292,
XX  PRO327, PRO1265, PRO344, PRO343, PRO347, PRO357, PRO715, PRO1017,
XX  PRO1112, PRO509, PRO853 and PRO882. These genes are amplified in
XX  the genome of tumour cells. The polypeptides are believed to contribute
XX  to tumourigenesis. The polypeptides are useful target for the
XX  identification of certain cancers, and may act as predictors of the
XX  prognosis of tumour treatment. Antibodies against these polypeptides
XX  are useful in the treatment and diagnosis of neoplastic cell growth
XX  and proliferation in mammals.
XX
XX  Sequence 317 AA;
XX
XX  Query Match      100.0%; Score 1720; DB 21; Length 317;
XX  Best Local Similarity 100.0%; Pred. No. 3.2e-147;
XX  Matches 317; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX  QY 1 MWVSGAPPALGGGCLGTFSTLLASTAILNAARIPVPPACGKPKQLNRVVGEDSTDSE 60
XX  DB 1 mvvsgappalgggclgtftslilastailnaaripvpacgkpkqlnrvvgedstdse 60
XX
XX  QY 61 WPWIVSTQKNGTHHCAGSLTSTRWVITAACFKONLNKPYLFSVLLGAWQLGNPGSRSQK 120
XX  DB 61 wpwivstqkngthhcagsltsrwrwvitaahcfknlkpylfsvllgawqlgnpgsrsk 120
XX
XX  QY 121 VGVAWVPHVPVYSWKEGACADIALVRLERSIQFSERVLPICLPDASIHLPNTHCWISGW 180
XX  DB 121 vgvaawvphvpvyswkegacadiavrlersiqfservlpiclpdasihlpnthcwisgw 180
XX
XX  QY 181 GSTQDGVPLPHPQTLQKLKVPIDSEVCSHLYWRGAGQGPITEDMLCAGYLEGERDACL 240
XX  DB 181 gsiqdgvpplphpqtlqlklkvpildsevcshlywrgagggpitedmlcagylegerdaclg 240
XX
XX  QY 241 DSGGPLMCQVDGAWLLAGIISWEGGCAERNRPGVYISLAHRSWVERKIVQGVQLRGAQG 300
XX  DB 241 dsggplmcqvdbgawllagiisweggcaernrpgvyislahrswwerkivqgvqlrgragq 300
XX
XX  QY 301 GGALRAPSGGGAARS 317
XX  DB 301 ggalrapsgsggaars 317
XX
XX  RESULT 5
XX  AA972890
XX  ID  AA972890 standard; Protein; 317 AA.
XX  XX
XX  AC  AA972890;
XX  XX
XX  DT 31-MAY-2001 (first entry)
XX  XX
XX  DE Human serine protease, protease C-E.
XX  XX
XX  KW Human; serine protease; protease C-E; therapy; desquamation; skin care;
XX  laundry detergent; shampoo; cleaning agent; hair care; skin flaking;
XX  neurodegenerative disorder; dermatological; immunogenic; proteolytic.
XX  chromosome 16p13.3.
XX  XX
XX  OS Homo sapiens.
XX  XX
XX  PN WO200116288-A2.
XX  XX
XX  PD 08-MAR-2001.
XX  XX
XX  PF 14-AUG-2000; 2000WO-US22117.
XX  XX
XX  PR 31-AUG-1999; 99US-0386529.

```

XX (ORTH ) ORTHO-MCNEIL PHARM INC.  
 XX Darrow A, Qi J, Andrade-Gordon P;  
 XX WPI: 2001-226681/23.  
 DR N-PSDB; AAD02990.  
 XX Novel serine protease termed protease C-E, useful for treating and  
 PT preventing skin flaking or imbalance of desquamation -  
 XX Claim 11; Fig 1; 78pp; English.  
 XX The present sequence is a human serine protease, protease C-E which  
 CC is a member of the S1 serine protease family. Protease C-E gene is  
 CC located on chromosome 16p13.3 and is expressed in pancreas, placenta,  
 CC prostate, small intestine, stomach, spleen, fibroblasts, epidermis,  
 CC cerebellum, cerebral cortex, pituitary and hippocampus. Protease C-E is  
 CC useful for treating an imbalance of desquamation, by topical application.  
 CC A non-pharmaceutical composition comprising the protein may be formulated  
 CC as a laundry detergent, shampoo, hard surface cleaning composition, dish  
 CC care cleaning composition, skin care composition and hair care  
 CC composition. Protease C-E is useful for treating and preventing skin  
 CC flaking, neurodegenerative disorders and dermatological pathologies. It  
 CC is less immunogenic to sensitive individuals and it provides efficient  
 CC proteolytic activity in a non-natural environment.  
 XX SQ Sequence 317 AA;

Query Match 100.0%; Score 1720; DB 22; Length 317;  
 Best Local Similarity 100.0%; Pred. No. 3.2e-147;  
 Matches 317; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVVSGAPPALGGGCLGTTFTSLLLASTAILNAARIPVPACGKPKQLNRVVGEDSTDSE 60  
 DB 1 mvvsgappalgggclgtftstlllastailnaaripvpacgkpkqlnrvvgedstdse 60  
 QY 61 WPWIVSTQKNGTHHCASLLTSRWVITAAHCFKDNLNKPYLFSVLLGAWQLNGPGRSQK 120  
 DB 61 wpwivstqkngthhcaglltsrwwitaahcfkdnlnkpylfsvllgawqlngpgrsqk 120  
 QY 121 VGVAWEPPHPVYSWKEGACADIALVRLERSIQFSERVLPICLPDASIHLPNTHCWISGW 180  
 DB 121 vgvaewephpvyswkegacadiavrlersiqfservlpiclpdasihlpnthcwlsigw 180  
 QY 181 GSIQDGVPLPHPOTLQKLKVPIDSEVCSHLYWRGAGQGPTEDMLCAGYLEGERDACLQ 240  
 DB 181 gsiqdgvpplphpptlqlkvpildsevcshlywrgagggptedmlcagylegerdaclg 240  
 QY 241 DSGGPLMCOVDGAWLLAGIISWGEGCAERNRPGVYISLSAHSRWKIVQGVQLRGAQG 300  
 DB 241 dsggplmcqvdgawllagilswgegcaernrpgvyislsahrsrwkivqgvqlrgragq 300  
 QY 301 GGALRAPSQSGGAARS 317  
 DB 301 ggalrapsgsggaars 317

RESULT 6  
 AAB80259  
 ID AAB80259 standard; Protein; 317 AA.  
 XX  
 AC AAB80259;  
 XX  
 DT 24-APR-2001 (first entry)  
 XX  
 DE Human PRO343 protein.  
 XX Human; PRO; dermatological; antipsoriatic; cytostatic; antiinflammatory;  
 KW antiparkinsonian nootropic; neuroprotective; vulnerary; cardiac;  
 KW antiangiogenic; vasotropic; antiasthmatic; antirheumatic; cancer;  
 KW antiarthritic; antiinfertility; antidiabetic; antiviral; diabetes;

XX ophthalmological; gene therapy; skin disease; gastrointestinal disorder;  
 KW ischaemia; inflammation.  
 XX Homo sapiens.  
 XX WO200104311-A1.  
 XX 18-JAN-2001.  
 XX 22-FEB-2000; 2000WO-US04414.  
 XX 07-JUL-1999; 99US-0143048.  
 XX 26-JUL-1999; 99US-0145698.  
 XX 28-JUL-1999; 99US-0146222.  
 XX 08-SEP-1999; 99WO-US20594.  
 XX 13-SEP-1999; 99WO-US20944.  
 XX 15-SEP-1999; 99WO-US21090.  
 XX 15-SEP-1999; 99WO-US21547.  
 XX 03-OCT-1999; 99WO-US23089.  
 XX 23-NOV-1999; 99WO-US28214.  
 XX 30-NOV-1999; 99WO-US28313.  
 XX 16-DEC-1999; 99WO-US30095.  
 XX 20-DEC-1999; 99WO-US30911.  
 XX 20-DEC-1999; 99WO-US30999.  
 XX 03-JAN-2000; 99WO-US00219.  
 XX (GETH ) GENENTECH INC.  
 XX Ashkenazi AJ, Botstein D, Desnoyers L, Eaton DL, Ferrara N;  
 PI Filvaroff E, Fong S, Gao W, Gerber H, Gerritsen ME, Goddard A;  
 PI Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ, Kljavin IJ;  
 PI Mather JP, Pan J, Paoni NF, Roy MA, Stewart TA, Tumas D;  
 PI Williams PM, Wood WI;  
 XX WPI: 2001-081051/09.  
 DR N-PSDB; AAF72420.  
 XX Sixty one nucleic acids encoding PRO polypeptides which are useful in  
 PT the treatment of skin diseases (e.g. psoriasis), cancers (e.g. lung  
 PT squamous cell carcinoma) and neurodegenerative diseases (e.g.  
 PT Alzheimer's disease) -  
 XX Claim 1; Fig 98; 393pp; English.  
 XX The present sequence is one of sixty one novel secreted and  
 CC transmembrane PRO polypeptides. The PRO polypeptides are  
 CC useful for treating skin diseases (e.g. psoriasis), cancers (e.g. lung  
 CC squamous cell carcinoma), gastrointestinal disorders (e.g.  
 CC enterocolitis), neurodegenerative diseases (e.g. Alzheimer's disease,  
 CC Parkinson's disease), wound repair, cardiovascular disorders (e.g.  
 CC endometrial bleeding angiogenesis, ischaemias such as coronary  
 CC ischaemia, atherosclerosis), inflammatory disorders (e.g. asthma,  
 CC rheumatoid arthritis, multiple sclerosis), infertility, AIDS and  
 CC diabetes and retinal disorders such as retinitis pigmentosa.  
 CC The PRO nucleic acids have applications in molecular biology, including  
 CC use as hybridization probes, and in chromosome and gene mapping.  
 XX Sequence 317 AA;

Query Match 100.0%; Score 1720; DB 22; Length 317;  
 Best Local Similarity 100.0%; Pred. No. 3.2e-147;  
 Matches 317; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVVSGAPPALGGGCLGTTFTSLLLASTAILNAARIPVPACGKPKQLNRVVGEDSTDSE 60  
 DB 1 mvvsgappalgggclgtftstlllastailnaaripvpacgkpkqlnrvvgedstdse 60  
 QY 61 WPWIVSTQKNGTHHCASLLTSRWVITAAHCFKDNLNKPYLFSVLLGAWQLNGPGRSQK 120  
 DB 61 wpwivstqkngthhcaglltsrwwitaahcfkdnlnkpylfsvllgawqlngpgrsqk 120  
 QY 121 VGVAWEPPHPVYSWKEGACADIALVRLERSIQFSERVLPICLPDASIHLPNTHCWISGW 180





PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0232196.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234597.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.

PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 08-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 03-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-476222/51.

N-PSDB; AAS26942.

Novel polypeptides and polynucleotides useful as diagnostic reagents to  
diagnose diseases or disorders associated with aberrant expression or  
activity of polypeptides, for treating blood clotting disorder,  
haemophilia

Claim 11; SEQ ID No 278; 601pp; English.

The invention relates to isolated nucleic acid molecules and their  
encoded secreted proteins. The nucleic acids and proteins are used to  
prevent, treat or ameliorate a medical condition in e.g. humans, mice,  
rabbits, goats, horses, cats, dogs, chickens or sheep. They  
are also used in diagnosing a pathological condition or susceptibility  
to a pathological condition. Antibodies to the proteins can also  
be used in alleviating symptoms associated with the disorders and in  
diagnostic immunoassays e.g. radioimmunoassays or enzyme linked  
immunosorbant assays (ELISA). Disorders which are diagnosed or treated  
include autoimmune diseases e.g. rheumatoid arthritis,  
hyperproliferative disorders e.g. neoplasms of the breast or liver,  
cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders  
e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.  
Alzheimer's disease, infections caused by bacteria, viruses and fungi  
and ocular disorders e.g. corneal infection, and many other



CC	disorders listed in the specification. The polypeptides can also	
CC	be used to aid wound healing and epithelial cell proliferation, to	
CC	prevent skin aging due to sunburn, to maintain organs before	
CC	transplantation, for supporting cell culture of primary tissues, to	
CC	regenerate tissues and in chemotaxis. The polypeptides can also be used	
CC	as a food additive or preservative to increase or decrease storage	
CC	capabilities, fat content, lipid, protein, carbohydrate, vitamins,	
CC	minerals, cofactors and other nutritional components. The present	
<hr/>		
	Query Match 100.0%; Score 1720; DB 22; Length 325;	
	Best Local Similarity 100.0%; Pred. No. 3.3e-147;	
	Matches 317; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
Qy	1 MVSAPPALGGCGTGTFTSLLLASTAILNAARIPVPACGKPOQLNRVYGGEDSTDSE 60	
Db		
Qy	61 WPWIVSIQKNGTHHCAGSLLSRWVITAACHCFKDNLNKPYLFSVLLGAWQLGNFGSRSQK 120	
Db		
Qy	121 VCVAVVEHPVYSWKEGACADIALVRLRSIQFSERVLPICLPDASIHLPNTHCWISGW 180	
Db		
Qy	181 GSIOQVPLPQTPQTLKLPITIDSEVCSHLYWRGAGOGPTEDMLCAGYLEGERDACLG 240	
Db		
Qy	241 DSGGLMCDVQAWLLAGIISWEGCAERNRPGVYISLSAHSRWVEKIVQGVQLRGAQG 300	
Db		
Qy	301 GGLRAPSGSGAARS 317	
Db		
<hr/>		
RESULT	9	
ID	AAU16966 standard; Protein; 351 AA.	
AC	AAU16966;	
DT	07-NOV-2001 (first entry)	
DE	Human novel secreted protein, SEQ ID 207.	
XX	Human; immunosuppressive; antiarthritic; antirheumatic;	
KW	cytostatic; cardiant; vasotropic; cerebroprotective; nootropic;	
KW	neuroprotective; antibacterial; virucide; fungicide; ophthalmological;	
KW	vulnary; secreted protein; rheumatoid arthritis;	
KW	hyperproliferative disorder; cardiovascular disorder; cardiac arrest;	
KW	cerebrovascular disorder; cerebral ischaemia; angiogenesis;	
KW	nervous system disorder; Alzheimer's disease; infection; ocular disorder;	
KW	corneal infection; wound healing; epithelial cell proliferation;	
XX	skin ageing; food additive; preservative; antiproliferative.	
OS	Homo sapiens.	
PN	WO200155441-A2.	
XX		
PD	02-AUG-2001.	
XX		
PF	17-JAN-2001; 2001WO-US01320.	
XX		
PR	31-JAN-2000; 2000US-0179065.	
PR	04-FEB-2000; 2000US-0180628.	
PR	24-FEB-2000; 2000US-0184664.	
PR	02-MAR-2000; 2000US-0186350.	
PR	16-MAR-2000; 2000US-0189874.	
PR	17-MAR-2000; 2000US-0190076.	
PR	18-APR-2000; 2000US-0198123.	

PR	19-MAY-2000; 2000US-0205515.	
PR	07-JUN-2000; 2000US-0209467.	
PR	28-JUN-2000; 2000US-0214886.	
PR	30-JUN-2000; 2000US-0215135.	
PR	07-JUL-2000; 2000US-0216647.	
PR	07-JUL-2000; 2000US-0216880.	
PR	11-JUL-2000; 2000US-0217487.	
PR	11-JUL-2000; 2000US-0217496.	
PR	14-JUL-2000; 2000US-0218290.	
PR	26-JUL-2000; 2000US-0220963.	
PR	26-JUL-2000; 2000US-0220964.	
PR	14-AUG-2000; 2000US-0224518.	
PR	14-AUG-2000; 2000US-0225213.	
PR	14-AUG-2000; 2000US-0225214.	
PR	14-AUG-2000; 2000US-0225266.	
PR	14-AUG-2000; 2000US-0225267.	
PR	14-AUG-2000; 2000US-0225268.	
PR	14-AUG-2000; 2000US-0225270.	
PR	14-AUG-2000; 2000US-0225447.	
PR	14-AUG-2000; 2000US-0225757.	
PR	14-AUG-2000; 2000US-0225758.	
PR	14-AUG-2000; 2000US-0225759.	
PR	18-AUG-2000; 2000US-0226279.	
PR	22-AUG-2000; 2000US-0226681.	
PR	22-AUG-2000; 2000US-0226686.	
PR	22-AUG-2000; 2000US-0227182.	
PR	23-AUG-2000; 2000US-0227009.	
PR	30-AUG-2000; 2000US-0228924.	
PR	01-SEP-2000; 2000US-0229287.	
PR	01-SEP-2000; 2000US-0229343.	
PR	01-SEP-2000; 2000US-0229344.	
PR	01-SEP-2000; 2000US-0229345.	
PR	05-SEP-2000; 2000US-0229509.	
PR	05-SEP-2000; 2000US-0229513.	
PR	06-SEP-2000; 2000US-0230437.	
PR	06-SEP-2000; 2000US-0230438.	
PR	08-SEP-2000; 2000US-0231242.	
PR	08-SEP-2000; 2000US-0231243.	
PR	08-SEP-2000; 2000US-0231244.	
PR	08-SEP-2000; 2000US-0231413.	
PR	08-SEP-2000; 2000US-0231414.	
PR	08-SEP-2000; 2000US-0232080.	
PR	08-SEP-2000; 2000US-0232081.	
PR	12-SEP-2000; 2000US-0231968.	
PR	14-SEP-2000; 2000US-0232397.	
PR	14-SEP-2000; 2000US-0232398.	
PR	14-SEP-2000; 2000US-0232399.	
PR	14-SEP-2000; 2000US-0232400.	
PR	14-SEP-2000; 2000US-0232401.	
PR	14-SEP-2000; 2000US-0233063.	
PR	14-SEP-2000; 2000US-0233064.	
PR	14-SEP-2000; 2000US-0233065.	
PR	21-SEP-2000; 2000US-0234223.	
PR	21-SEP-2000; 2000US-0234274.	
PR	25-SEP-2000; 2000US-0234997.	
PR	25-SEP-2000; 2000US-0234998.	
PR	26-SEP-2000; 2000US-0235484.	
PR	27-SEP-2000; 2000US-0235834.	
PR	27-SEP-2000; 2000US-0235836.	
PR	29-SEP-2000; 2000US-0236327.	
PR	29-SEP-2000; 2000US-0236367.	
PR	29-SEP-2000; 2000US-0236368.	
PR	29-SEP-2000; 2000US-0236369.	
PR	29-SEP-2000; 2000US-0236370.	
PR	02-OCT-2000; 2000US-0236802.	
PR	02-OCT-2000; 2000US-0237037.	
PR	02-OCT-2000; 2000US-0237038.	
PR	02-OCT-2000; 2000US-0237039.	
PR	02-OCT-2000; 2000US-0237040.	
PR	13-OCT-2000; 2000US-0239935.	
PR	13-OCT-2000; 2000US-0239937.	
PR	20-OCT-2000; 2000US-0240960.	

20-OCT-2000; 2000US-0241221.  
 20-OCT-2000; 2000US-0241785.  
 20-OCT-2000; 2000US-0241786.  
 20-OCT-2000; 2000US-0241787.  
 20-OCT-2000; 2000US-0241808.  
 20-OCT-2000; 2000US-0241809.  
 20-OCT-2000; 2000US-0241826.  
 01-NOV-2000; 2000US-0244617.  
 08-NOV-2000; 2000US-0246474.  
 08-NOV-2000; 2000US-0246475.  
 08-NOV-2000; 2000US-0246476.  
 08-NOV-2000; 2000US-0246477.  
 08-NOV-2000; 2000US-0246478.  
 08-NOV-2000; 2000US-0246523.  
 08-NOV-2000; 2000US-0246524.  
 08-NOV-2000; 2000US-0246525.  
 08-NOV-2000; 2000US-0246526.  
 08-NOV-2000; 2000US-0246527.  
 08-NOV-2000; 2000US-0246528.  
 08-NOV-2000; 2000US-0246532.  
 08-NOV-2000; 2000US-0246609.  
 08-NOV-2000; 2000US-0246610.  
 08-NOV-2000; 2000US-0246611.  
 08-NOV-2000; 2000US-0246613.  
 17-NOV-2000; 2000US-0249207.  
 17-NOV-2000; 2000US-0249208.  
 17-NOV-2000; 2000US-0249209.  
 17-NOV-2000; 2000US-0249210.  
 17-NOV-2000; 2000US-0249211.  
 17-NOV-2000; 2000US-0249212.  
 17-NOV-2000; 2000US-0249213.  
 17-NOV-2000; 2000US-0249214.  
 17-NOV-2000; 2000US-0249215.  
 17-NOV-2000; 2000US-0249216.  
 17-NOV-2000; 2000US-0249217.  
 17-NOV-2000; 2000US-0249218.  
 17-NOV-2000; 2000US-0249219.  
 17-NOV-2000; 2000US-0249244.  
 17-NOV-2000; 2000US-0249245.  
 17-NOV-2000; 2000US-0249264.  
 17-NOV-2000; 2000US-0249265.  
 17-NOV-2000; 2000US-0249297.  
 17-NOV-2000; 2000US-0249299.  
 17-NOV-2000; 2000US-0249300.  
 01-DEC-2000; 2000US-0250160.  
 01-DEC-2000; 2000US-0250391.  
 05-DEC-2000; 2000US-0251030.  
 05-DEC-2000; 2000US-0251988.  
 06-DEC-2000; 2000US-0251479.  
 08-DEC-2000; 2000US-0251856.  
 08-DEC-2000; 2000US-0251868.  
 08-DEC-2000; 2000US-0251869.  
 08-DEC-2000; 2000US-0251989.  
 08-DEC-2000; 2000US-0251990.  
 11-DEC-2000; 2000US-0254097.  
 05-JAN-2001; 2001US-0259678.  
 (HUMA-) HUMAN GENOME SCI INC.  
 Rosen CA, Barash SC, Ruben SM;  
 WPI: 2001-476222/51.  
 N-PSDB: AAS26871.  
 Novel polypeptides and polynucleotides useful as diagnostic reagents to  
 diagnose diseases or disorders associated with aberrant expression or  
 activity of polypeptides, for treating blood clotting disorder,  
 hemophilia  
 Claim 11; SEQ ID NO 207; 601pp; English.  
 The invention relates to isolated nucleic acid molecules and their  
 encoded secreted proteins. The nucleic acids and proteins are used to

prevent, treat or ameliorate a medical condition in e.g. humans, mice,  
 rabbits, goats, horses, cats, dogs, chickens or sheep. They  
 are also used in diagnosing a pathological condition or susceptibility  
 to a pathological condition. Antibodies to the proteins can also  
 be used in alleviating symptoms associated with the disorders and in  
 diagnostic immunoassays e.g. radioimmunoassays or enzyme linked  
 immunosorbant assays (ELISA). Disorders which are diagnosed or treated  
 include autoimmune diseases e.g. rheumatoid arthritis,  
 hyperproliferative disorders e.g. neoplasms of the breast or liver,  
 cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders  
 e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.  
 Alzheimer's disease, infections caused by bacteria, viruses and fungi  
 and ocular disorders e.g. corneal infection, and many other  
 disorders listed in the specification. The polypeptides can also  
 be used to aid wound healing and epithelial cell proliferation, to  
 prevent skin aging due to sunburn, to maintain organs before  
 transplantation, for supporting cell culture of primary tissues, to  
 regenerate tissues and in chemotaxis. The polypeptides can also be used  
 as a food additive or preservative to increase or decrease storage  
 capabilities, fat content, lipid, protein, carbohydrate, vitamins,  
 minerals, cofactors and other nutritional components. The present

Query Match 100.0%; Score 1720; DB 22; Length 351;  
 Best Local Similarity 100.0%; Pred. No. 3.6e-147; Indels 0; Gaps 0;  
 Matches 317; Conservative 0; Mismatches 0;

Qy 1 MVSGAPPALGGGCGTFTSLLLASTAILNAARIPVPACGKQQLNRVVGSDTDE 60  
 Dd 35 mvsgappalggcggtftslillastailnaaripvpacgkqqlnrvggedstdse 94  
 Qy 61 WPWIVSTQKNGTHHCAGSLTSTRWVITAAHCFKONLKNPYLFLSVLLGAWQLGNPGRSQK 120  
 Dd 95 wpwivsiqkngthhcagslltsrwvitaahcfkdnlnkpylflsvllgawqlgnpgrsqk 154  
 Qy 121 VGVAVVEPHVPVYSNKEGACADIALVRLERSIOFSERVLPICLPDASHLPPNTHCWISGW 180  
 Dd 155 vgvavvephvpvyswkegacacialvrlersiqfservlpiclpdaslnpntncwlsigw 214  
 Qy 181 GSIQDGVPLPHPQTLQKLVPIIDSEVCSHLYWRGAGGPITEDMLCAGYLEGRDACLG 240  
 Dd 215 gsiqdgvpplphpqtlqlkvpildsevcshlywrgaggpitedmlcagylegrdaclg 274  
 Qy 241 DSGGFLMCOVDGANLLAGIISWEGGCAERNRPVYISLSAHSRWEKIVQVQLRGAQG 300  
 Dd 275 dsggplmcqvqdgawllagilswgegcarnrpvgvylsahrswekivqvgqlrgragq 334  
 Qy 301 GGALRAPSQSGAAARS 317  
 Dd 335 ggalrapsgsgaaars 351

RESULT 10  
 AAU23751  
 ID AAU23751 standard; Protein; 322 AA.  
 XX  
 AC AAU23751;  
 XX  
 DT 18-DEC-2001 (first entry)  
 XX  
 DE Novel human enzyme polypeptide #837.  
 XX  
 KW Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;  
 KW ligase; hyperproliferative disorder; immunodeficiency disorder;  
 KW autoimmune disorder; neurological disorder; metabolic disorder;  
 KW inflammatory disorder; cardiovascular disorder; reproductive disorder;  
 KW blood-related disorder; infectious disorder; cytostatic; anti arthritic;  
 KW nephrotropic; anticoagulant.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200155301-A2.  
 XX  
 PD 02-AUG-2001.

XX 17-JAN-2001; 2001WO-US01239.  
PF  
XX 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214986.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226686.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233053.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX Rosen CA, Barash SC, Ruben SM;  
XX WPI; 2001-465566/50.  
XX N-PSDB; AAS41621.  
DR



```
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-024617.
PR 08-NOV-2000; 2000US-0246174.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 03-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.

PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX
XX WPI: 2001-476222/51.
XX N-PSDB; AAS26943.
XX
XX Novel polypeptides and polynucleotides useful as diagnostic reagents to
XX diagnose diseases or disorders associated with aberrant expression or
XX activity of polypeptides, for treating blood clotting disorder,
XX haemophilia
XX
XX Claim 11; SEQ ID No 279; 601pp; English.
XX
XX The invention relates to isolated nucleic acid molecules and their
XX encoded secreted proteins. The nucleic acids and proteins are used to
XX prevent, treat or ameliorate a medical condition in e.g. humans, mice,
XX rabbits, goats, horses, cats, dogs, chickens or sheep. They
XX are also used in diagnosing a pathological condition or susceptibility
XX to a pathological condition. Antibodies to the proteins can also
XX be used in alleviating symptoms associated with the disorders and in
XX diagnostic immunoassays e.g. radioimmunoassays or enzyme linked
XX immunosorbant assays (ELISA). Disorders which are diagnosed or treated
XX include autoimmune diseases e.g. rheumatoid arthritis,
XX hyperproliferative disorders e.g. neoplasms of the breast or liver,
XX cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
XX e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.
XX Alzheimer's disease, infections caused by bacteria, viruses and fungi
XX and ocular disorders e.g. corneal infection, and many other
XX disorders listed in the specification. The polypeptides can also
XX be used to aid wound healing and epithelial cell proliferation, to
XX prevent skin aging due to sunburn, to maintain organs before
XX transplantation, for supporting cell culture of primary tissues, to
XX regenerate tissues and in chemotaxis. The polypeptides can also be used
XX as a food additive or preservative to increase or decrease storage
XX capabilities, fat content, lipid, protein, carbohydrate, vitamins,
XX minerals, cofactors and other nutritional components. The present
XX
XX Query Match 99.7%; Score 1714; DB 22; Length 322;
XX Best Local Similarity 99.7%; Pred. No. 1.1e-146;
XX Matches 316; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MVVSGAPPALGGGCLGTTFTSLLLLASTAILNAARIPVPACGKPPQQLNRVVGGEDSTDSE 60
Db 6 mvvsgappaigggclgtftstllllastailnaaripvpacgkppqqlnrvvvggedstdse 65
QY 61 WPWIVSIQKNGTHHCAGSLTTSRWVITAAHCFKDNLNKPYLFSVLGAWOLNPGRSRQK 120
Db 66 wpwivsixkngthhcagslltstwtiaahcfkdnlnkpylfsvlglawolnpgrsrsk 125
QY 121 VGVAVVEPHVPVYSWKEGACADIALVRLERSIQFSERVLPICLPDASIHLPNTHCHWISGW 180
Db 126 vgvavvephvpvyswkegacadiavrlersiqfservlpiclpdasihlpnthchwisgw 185
QY 181 GSTODGVPLPHTQLOKLYPIIDSEVCSHLYWRGAGGPTEDMLCAGYLEGERDACLG 240
Db 186 gsiqdgvpplphpqtlqklkypiidsevcshlywrgagggptedmlcagylegerdacly 245
QY 241 DSGGPLMCQVDGAWLLAGIISWEGCAERNRPGVYISLSAHRSWVEKIVGQVLRGAAQ 300
Db 246 dsggplmcqvdgawllagiiswgegcaernrpgvylsahrswekivgqvlgrrgaqg 305
QY 301 GGALRAPSGSGAAARS 317
Db 306 ggalrapsgsgaaars 322
```



FT	Misc-difference	196	/note= "likely to be involved in disulphide bonding"
FT	Disulfide-bond	202	/note= "likely to be a catalytic residue"
FT	Disulfide-bond	220	/note= "likely to be involved in disulphide bonding"
FT	Misc-difference	235	/note= "likely to be involved in disulphide bonding"
XX	WO9836054-A1.		/note= "likely to be a catalytic residue"
PN	20-AUG-1998.		
XX	13-FEB-1998;	98WO-AU000085.	
XX	18-NOV-1997;	97AU-0000422.	
PR	13-FEB-1997;	97AU-0005101.	
XX	(AMRA-) AMRAD OPERATIONS PTY LTD.		
PA	Antalis TM, Hooper JD;		
XX	WPI; 1998-480768/41.		
DR	N-PSDB; AAV59134.		
XX	New serine protease(s) and kinase involved in regulating cell activity and viability - particularly the testis-specific protease HELA2 used for modulation of fertility and as tumour suppressor		
PT	Example 15; Fig 20A; 167pp; English.		
XX	AAW77302-04 represent HELA2 homologues. The genes are found in a cluster on chromosome 16p13.3. HELA2 was isolated from HeLa cells, and has homology to serine proteases. The protein is involved in or associated with regulation of cell activity and/or viability. Administration of recombinant HELA2 (also called testisin) is used to increase fertility. Downregulation of HELA2 reduces fertility. HELA2 is also a suppressor of testicular germ cell cancers (seminoma) and is also expressed in some non-testicular cancers (of colon, pancreas, prostate and ovary), so is a marker/potential therapeutic target for cancer. The promoter from the HELA2 gene is useful for testis-specific expression of other genes, e.g. for gene therapy or modulation of fertility. Drugs that block activity of HELA2 should have antitumour activity (other than in testis) while in testis recombinant HELA2 should stop growth of tumours and normalise sperm development (eliminating the need for orchidectomy). Identification of mutant forms of HELA2 can be used to diagnose infertility.		
XX	Sequence	271 AA;	
QY	Query Match	86.3%; Score 1484; DB 19; Length 271;	
QY	Best Local Similarity	100.0%; Pred. No. 5.5e-126;	
QY	Matches	271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	47	LNRVVGGEDSTDSEWPWTIVSIQKNGTHHCAGSLTTSRWITAAHCFKDNLNKPYLFSVLL 106	
Db	1	lnrvvggedstdsewpwivsiqkngthhcagslltstwitaahcfkdnlnkpylfsvll 60	
QY	107	GAWOLGNPGRSQRKGVAVVEPHFPVYSWKEGACADIALVRLERSIQFSERVLPICLPDAS 166	
Db	61	gawlgnpgrsqkqvavvephfpvyswkegacadiavrlersiqfservlpiclpdas 120	
QY	167	IHLPPNTHCWSIGWSGSTDQGVPLPHPQTQKLKVPIDIDSEVCSHLYNRGAGCGPITEDML 226	
Db	121	ihlppnthcwsigwsgiqdgvplphpqtqlklkvpdiidsevcshlywrgagcgpitedml 180	
QY	227	CAGVLEGERDACLDGSGPLMCQVDGAWLLAGIISWEGGCAERNRPGVYISLSAHRSWVE 286	
Db	181	cagvlegerdacldgsggplmcqvdgawllagiisweggcaernrpgvyislsahrswve 240	
QY	287	KIVOGVOLIRGAOGGGALRAPSGSGAAARS 317	





---

THIS PAGE BLANK (USPTO)

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2002, 10:12:26 ; Search time 12.99 seconds  
(without alignments)  
596.067 Million cell updates/sec

Title: US-09-903-925a-263  
Perfect score: 1720  
Sequence: 1 MVVSGAPALGGCLGTFTS.....AOGGALRAPSGSGAARS 317

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*

- 1: /cgn2\_6/ptodata/2/iaa/5A.COMB.pep.\*
- 2: /cgn2\_6/ptodata/2/iaa/5B.COMB.pep.\*
- 3: /cgn2\_6/ptodata/2/iaa/6A.COMB.pep.\*
- 4: /cgn2\_6/ptodata/2/iaa/6B.COMB.pep.\*
- 5: /cgn2\_6/ptodata/2/iaa/PTUS.COMB.pep.\*
- 6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	581.5	33.8	299	4	US-08-944-483-66
2	554	32.2	276	2	US-09-016-366A-15
3	554	32.2	276	2	US-08-978-404B-21
4	538.5	31.3	274	2	US-08-978-404B-5
5	531.5	30.9	314	4	US-09-008-271A-3
6	530.5	30.8	274	2	US-09-016-366A-21
7	530.5	30.8	274	2	US-08-978-404B-16
8	530.5	30.8	275	2	US-09-016-366A-17
9	530.5	30.8	275	2	US-08-978-404B-12
10	529.5	30.8	273	2	US-09-016-366A-19
11	529.5	30.8	273	2	US-08-978-404B-14
12	522	30.3	270	2	US-08-978-404B-8
13	519.5	30.2	273	2	US-08-978-404B-6
14	518	30.1	273	2	US-08-978-404B-3
15	514.5	29.9	267	2	US-09-016-366A-23
16	514.5	29.9	267	2	US-08-978-404B-18
17	511.5	29.7	245	4	US-08-944-483-69
18	511.5	29.7	249	4	US-09-079-970A-5
19	508	29.5	304	4	US-09-088-651-2
20	506.5	29.4	245	4	US-09-079-970A-6
21	473.5	27.5	248	4	US-08-944-483-63
22	470.5	27.4	238	4	US-08-944-483-64
23	469.5	27.3	638	2	US-08-681-151-3
24	468.5	27.2	435	4	US-09-008-271A-6
25	466	27.1	454	4	US-09-518-046-2
26	463	26.9	416	2	US-09-000-846-2
27	462	26.9	492	4	US-09-342-749-2

28	453	26.3	283	3	US-08-807-151-1	Sequence 1, Appli
29	453	26.3	283	4	US-09-478-957-1	Sequence 1, Appli
30	449	26.1	791	1	US-08-643-219-1	Sequence 1, Appli
31	449	26.1	791	3	US-08-851-350-1	Sequence 1, Appli
32	447	26.0	814	1	US-08-750-711-1	Sequence 1, Appli
33	446	25.9	546	6	5200340-6	Patent No. 5200340
34	446	25.9	790	1	US-08-469-486-54	Sequence 54, Appl
35	446	25.9	790	2	US-08-469-658-54	Sequence 54, Appl
36	446	25.9	791	2	US-09-131-995-1	Sequence 1, Appli
37	446	25.9	791	2	US-08-832-087B-1	Sequence 1, Appli
38	446	25.9	791	4	US-09-132-154-1	Sequence 1, Appli
39	446	25.9	810	1	US-07-854-603-2	Sequence 2, Appli
40	446	25.9	810	1	US-08-147-000B-29	Sequence 29, Appli
41	446	25.9	810	4	US-09-086-514-1	Sequence 1, Appli
42	446	25.9	810	6	5200340-8	Patent No. 5200340
43	445.5	25.9	256	2	US-09-027-337-3	Sequence 3, Appli
44	443.5	25.8	798	1	US-08-200-900A-2	Sequence 2, Appli
45	443.5	25.8	798	5	PCT-US94-00616-2	Sequence 2, Appli

#### ALIGNMENTS

RESULT 1  
US-08-944-483-66  
; Sequence 66..Application US/08944483  
; Patent No. 6232456

; GENERAL INFORMATION:

; APPLICANT: COHEN, MAURICE

; APPLICANT: COLPITTS, TRACEY L.

; APPLICANT: FRIEDMAN, PAULA N.

; APPLICANT: GRANADOS, EDWARD N.

; APPLICANT: KLASS, MICHAEL R.

; APPLICANT: RUSSELL, JOHN C.

; APPLICANT: STEWART, KENT D.

; APPLICANT: STROUPE, STEVEN D.

; TITLE OF INVENTION: NOVEL SERINE PROTEASE REAGENTS

; TITLE OF INVENTION: AND METHODS USEFUL FOR DETECTING AND TREATING DISEASES

; NUMBER OF SEQUENCES: 76

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Abbott Laboratories

; STREET: 100 Abbott Park Road

; CITY: Abbott Park

; STATE: IL

; COUNTRY: USA

; ZIP: 60064-3500

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FASTSEQ for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/944,483

; FILING DATE:

; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Becker, Cheryl L.

; REGISTRATION NUMBER: 35,441

; REFERENCE/DOCKET NUMBER: 6183.US.01

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 847/935-1729

; TELEFAX: 847/938-2623

; TELEX:

; INFORMATION FOR SEQ ID NO: 66:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 299 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

MOLECULE TYPE: No. 6232456e  
US-08-944-483-66

Query Match 33.8%; Score 581.5; DB 4; Length 299;  
Best Local Similarity 45.4%; Pred. No. 6.5e-50;  
Matches 113; Conservative 42; Mismatches 89; Indels 5; Gaps 4;  
QY 50 VVGEDSTDEWPIVSIQKNGTHHCAGSLTTSRWITAAHCFKDNKPYLFVSVLLGAW 109  
DB 1 ITGSSAVACQWPHQVSIYEGVHVCGLVSEQWVLSAAHCFSEHKE-AYEVKLGAAH 59  
QY 110 QLNPGSRQKVGVAWPEHPVPSWKEGACADIALVRLERSIQFSERVLPICLPDASIH 169  
DB 60 QLDSEDAKVSTLKDIIHPSY-LOEGSQDIALQLSRPITFSRVRIPICLPAANASE 118  
QY 170 PPNTHCWISGWSIQDGVPLPHQTLQKLVPIIDSEVCSHLYWRGA-GQGP--ITEDML 226  
DB 119 PNGLHCVTGTGHWVAPSVSLTPKPLQOLEVPLISRETCLYNIDAKPEEPHFVQEDMV 178  
QY 227 CAGYLEGERDACLDGSGGLMCOVDGAWLLAGIISWEGCAERNRPGVWISLSAHRSWVE 286  
DB 179 CAGIVEGKDACQDGSGLPCPVEGLWYLTGIYVSWGDACGARNRPGVYTLASSYASW 238  
QY 287 KIVGVOLR 295  
DB 239 SKVTELQPR 247

RESULT 2  
US-09-016-366A-15  
; Sequence 15, Application US/09016366A  
; Patent No. 5955431  
; GENERAL INFORMATION:  
; APPLICANT: Stevens, Richard L.  
; APPLICANT: Huang, Chifu  
; TITLE OF INVENTION: MAST CELL PROTEASE PEPTIDE  
; NUMBER OF INVENTIONS: INHIBITORS  
; NUMBER OF SEQUENCES: 65  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.  
; STREET: 600 Atlantic Avenue  
; CITY: Boston  
; STATE: MA  
; COUNTRY: U.S.A.  
; ZIP: 02210-2211  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/016,366A  
; FILING DATE: January 30, 1998  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/037,090  
; FILING DATE: 05-FEB-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Plumer, Elizabeth R.  
; REGISTRATION NUMBER: 36,637  
; REFERENCE/DOCKET NUMBER: B0801/7093  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617-720-3500  
; TELEFAX: 617-720-2441  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 15:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 276 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein

US-09-016-366A-15

Query Match 32.2%; Score 554; DB 2; Length 276;  
Best Local Similarity 41.6%; Pred. No. 3.2e-47;  
Matches 116; Conservative 48; Mismatches 93; Indels 22; Gaps 9;  
QY 21 LLLIATAILNAARIPVP-PACGKPQOLNRVVGEDSTDEWPIVSIQ---KNGTHHCA 76  
DB 6 LLLWALSILASLYVSAPRA----NQRVIGVGGHEASESKWPQVSLRFLKLTWIFECG 61  
QY 77 GSLTTSRWITAAHCFKDNKPYLFVSVLLGAWOL--GNPGRSQKVGVAWPEHPVYSW 134  
DB 62 GSLTHPQWLTAACHVGHPIKSPOLFRVQLREQVLYYGD-----QLLSLNRIVVHPHYT 116  
QY 135 KEGACADIALVRLERSIQFSERVLPICLPDASIHLPNTHCWISGWSIQDGVPLPHQ 194  
DB 117 AEGG-ADVALLELEVNVNTHIHPISLPASETFPPGTSCWVTGWGDIDNDEPLPPYP 175  
QY 195 LQKLKVPIDSEVCSHLYWRGAGQ---PITED-MLCAGYLEGERDACLDGSGGLMCOV 250  
DB 176 LKQVKVPIVENSCLDRKYHTGLYTGDDFPPIVHDGMLCAG--NTRDSCQDSDGGLVCKV 233  
QY 251 DGAWLLAGIISWEGCAERNRPGVWISLSAHRSWVEKIV 289  
DB 234 KGTWLQAGVSVWEGCAQPNKPGIYTRVYLDWIHRYV 272

RESULT 3  
US-08-978-404B-21  
; Sequence 21, Application US/08978404B  
; Patent No. 5968782  
; GENERAL INFORMATION:  
; APPLICANT: Stevens, Richard L.  
; TITLE OF INVENTION: MAST CELL PROTEASE THAT CLEAVES  
; TITLE OF INVENTION: FIBRINOGEN  
; NUMBER OF SEQUENCES: 74  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.  
; STREET: 600 Atlantic Avenue  
; CITY: Boston  
; STATE: MA  
; COUNTRY: U.S.A.  
; ZIP: 02210-2211  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/978,404B  
; FILING DATE: 25-NOV-97  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/032,354  
; FILING DATE: 04-DEC-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Plumer, Elizabeth R.  
; REGISTRATION NUMBER: 36,637  
; REFERENCE/DOCKET NUMBER: B0801/7090  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617-720-3500  
; TELEFAX: 617-720-2441  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 21:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 276 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: No. 5968782e  
US-08-978-404B-21

Query Match	32.2%	Score 554;	DB 2;	Length 276;
Best Local Similarity	41.6%;	Pred. No. 3.2e-47;		
Matches 116;	Conservative 48;	Mismatches 93;	Indels 22;	Gaps
QY	21	LLLLASTAINLAARIPVP--PACGKPOOLNRVGGESTDSEWPWISIO---KNGTHFCA	76	
Db	6	LLLLWALSLLASLVYAPRA----NORCVIGGHEASSKVPQWVSLRFLKNYWHFCG	61	
QY	77	GSLLTSRWVITAAHCFKDNLKKPLSFVLLGAWQL--GNPGRSRSQKVGVAVPEHPVYSW	134	
Db	62	GSIIHPQWMLTAAHCVGHPIKSPQLFVRQLREQLYGD----QLLSNRIVVVHPHYIT	116	
QY	135	KEGACADIALVRLERSITQFSERVLPICLPDASHLHPNTHCWISGWSIQDQVGPLPHQPT	194	
Db	117	AEGG-ADVALLELEVPVNVSTHPIPLSPASETFPPGTSCWVTGWGIDINDPELPPPPY	175	
QY	195	LOKLKVPIDSEVCSHLYWRGAGG--PITED-MLCAGYLEGERDACIGDSGGPLMCQV	250	
Db	176	LKQVKVPVENSICDKRHYTGLTGTODDFPIVDHGMJLACG--NTRRDSOCODSGGPLVCKV	233	
QY	251	DGAWLLAGIISWGEGCAERNRPGVYISLSAHRSSVWEKIV	289	
Db	234	KGTWLOAGVYSWGEGCAOPNKPQIYTRVYIYLDWIHRV	272	

```

RESULT      4
US-08-978-404B-5
; Sequence 5, Application US/08978404B
; Patent No. 5968782
; GENERAL INFORMATION:
; APPLICANT: Stevens, Richard L.
; TITLE OF INVENTION: FAST CELL PROTEASE THAT CLEAVES
; TITLE OF INVENTION: FIBRINOGEN
; NUMBER OF SEQUENCES: 74
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: MA
; COUNTRY: U.S.A.
; ZIP: 02210-2211
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/978,404B
; FILING DATE: 25-NOV-97
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/032,354
; FILING DATE: 04-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Plumer, Elizabeth R.
; REGISTRATION NUMBER: 36,637
; REFERENCE/DOCKET NUMBER: B0801/7090
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-720-3500
; TELEFAX: 617-720-2441
; TELEX:
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 274 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: NO. 5968782e
US-08-978-404B-5

```

Query Match 31.3%; Score 538.5; DB 2; Length 274;

```

Best Local Similarity 40.6%; Pred. No. 1.1e-45;
Matches 117; Conservative 46; Mismatches 84; Indels 41; Gaps

QY 21 LLLLAST--AILNAARIPVPPACGKPKQQLNRVVGGEDSTDSEWPWIVSIQKNGT---HH 74
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 5 LLLLALSPLASIVHAAPCPVKQVGVG-----IVGGREASESKWPNQVSLRPFKFSWMHF 57
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :

QY 75 CAGSLTTSRWVITAAHCFKDNLANKPYLFVLLGAWOLGNPGRSSQ-----KVGVM 125
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 58 CGGSLHPQWLTAACHVGLHIKSPFLRVQL-----REQLYYADQLLLTVNRTV 107
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :

QY 126 VEPHPYWSKEGACADIALVRLERSIQFSERVLPICLPDASIHLPNTHCWSWGMSIOD 185
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 108 VHPH-YTVEDG--ADIALELEIPVNVSTHHPISLPFASFTFPGTSCWVTGMGDIDS 164
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :

QY 186 GVPLPQPTQLKLVKPIIDSEVCVSHLYWRGAGCG---PITD-MLCAGYLEGERDACLGD 241
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 165 DEPLPPYLPKQVKVPVENSCLDRKYHTCLYTGDDVPVQDGLCAG--NTRSDSCQGD 222
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :

QY 242 SGGPLMCQVDGAWLLAGITISWGEGCAERNRPGVVISLSAHRSWKXIV 289
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 223 SGGPLCVKVGKGTWLQAGVWSWGEGCAEARNRPGIYTRVTYLDWIHRV 270
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :

RESULT 5
US-09-008-271A-3
; Sequence 3, Application US/09008271A
; Patent No. 6203979
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; Hillman, Jennifer L.
; Yue, Henry
; Guegler, Karl J.
; Corley, Neil C.
; Tang, Tom Y.
; Shah, Purvi
;
; TITLE OF INVENTION: HUMAN PROTEASE MOLECULES
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Dr.
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/008,271A
; FILING DATE: 16-Jan-1998
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: <Unknown>
; FILING DATE: <Unknown>
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Mohan-Peterson, Sheela
; REGISTRATION NUMBER: 41,201
; REFERENCE/DOCKET NUMBER: PF-0458 US
;
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
;
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 314 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: PROSTUT03
; CLONE: 789927
; SEQUENCE DESCRIPTION: SEQ ID NO: 3 :

```

RESULT 6  
US-09-016-366A-21  
; Sequence 21, Application US/09016366A  
; Patent No. 5955431  
; GENERAL INFORMATION:  
; APPLICANT: Stevens, Richard L.  
; APPLICANT: Huang, Chifu  
; TITLE OF INVENTION: MAST CELL PROTEASE PEPTIDE  
; TITLE OF INVENTION: INHIBITORS  
; NUMBER OF SEQUENCES: 65  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.  
; STREET: 600 Atlantic Avenue  
; CITY: Boston  
; STATE: MA  
; COUNTRY: U.S.A.  
; ZIP: 02210-2211  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/016.366A  
; FILING DATE: January 30, 1998  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/037,090  
; FILING DATE: 05-FEB-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Plumer, Elizabeth R.  
; REGISTRATION NUMBER: 36,637  
; REFERENCE/DOCKET NUMBER: B0801/7093  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617-720-3500  
; TELEFAX: 617-720-2441  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 21:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 274 amino acids  
; TYPE: amino acid

```

RESULT 7
US-08-978-404B-16
; Sequence 16, Application US/08978404B
; Patent No. 5968782
; GENERAL INFORMATION:
; APPLICANT: Stevens, Richard L.
; TITLE OF INVENTION: MAST CELL PROTEASE THAT CLEAVES
; TITLE OF INVENTION: FIBRINOGEN
; NUMBER OF SEQUENCES: 74
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: MA
; COUNTRY: U.S.A.
; ZIP: 02210-2211
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/978,404B
; FILING DATE: 25-NOV-97
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/032,354
; FILING DATE: 04-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Plumer, Elizabeth R.
; REGISTRATION NUMBER: 36,637
; REFERENCE/DOCKET NUMBER: B0801/7090
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-720-3500
; TELEFAX: 617-720-2441
; TELEX:
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 274 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 5968782e

```

9:05

;  
US-09-016-366A-17

aps,

US-08-978-404B-12







Qy 78 SLTSRWVITAAHCFKDN-----LNKPYLFSVLLGAWLGNFGSRQSKYGVAMVE 127  
||: ||:||||| | | ||: : | :  
Db 60 SLIHPQWVLTAAHCVPNKPDLKRLQKQYLY-----YHDHLMTVSQII 106  
Qy 128 PHP-VYSWKEGACADIALVRLERSIQSERVLPICLPDASIHLPNTHCWISWGSIDQGV 187  
|| | :|| ||||| : : | || | | ||:||||| :  
Db 107 SHDPFYTAQDG--ADIALKLITNPVNI SDYVHPVLPPLPASETFPSGTLCTWGTGNGNIDNGV 165  
Qy 187 VPLHPOTLOKLVPIIDSEVCSHLYWRGAGQGP-----ITEDMLCAGYLEGERDACLGDSG 243  
|| | :|| ||||| : | : ||||| | | :|||  
Db 165 VSLPPPPLEVEQVPIVENRLCDLKYHKGLTGNVHVIRDDMLCAGN-EG-HDSCQGDG 223  
Qy 243 GGLMCQVDGAWLLAGIISWEGGCAERNRPGVYISLSAHSRWSWEKIV 289  
|||:|:| | ||:||||||| ||||| : : | :  
Db 223 GGLVCKVEDTWLQAGVYVSWEGGCAQPNRPGIYTRVYIYLDWIYRYV 269

## RESULT 14

US-08-978-404B-3  
; Sequence 3, Application US/08978404B  
; Patent No. 5968782  
; GENERAL INFORMATION:  
; APPLICANT: Stevens, Richard L.  
; TITLE OF INVENTION: MAST CELL PROTEASE THAT CLEAVES  
; TITLE OF INVENTION: FIBRINOGEN  
; NUMBER OF SEQUENCES: 74  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.  
; STREET: 600 Atlantic Avenue  
; CITY: Boston  
; STATE: MA  
; COUNTRY: U.S.A.  
; ZIP: 02210-2211  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FASTSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/978,404B  
; FILING DATE: 25-NOV-97  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/032,354  
; FILING DATE: 04-DEC-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Plumer, Elizabeth R.  
; REGISTRATION NUMBER: 36,637  
; REFERENCE/DOCKET NUMBER: B0801/7090  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617-720-3500  
; TELEFAX: 617-720-2441  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 273 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: No. 5968782e  
US-08-978-404B-3

Query Match 30.1%; Score 518; DB 2; Length 273;  
Best Local Similarity 38.5%; Pred. No. 1.2e-43;  
Matches 110; Conservative 53; Mismatches 85; Indels 38; Gaps 9;  
Qy 22 LLLASTAILNAARIPVPPACKPQOLNRVVGEDSTSEWPIVSIQKNET---HHCAGS 78  
|| | :|| | | :||| :||| ||| : | |||  
Db 4 LLLLTLLLSL---VHAAPGPMATREGIVGGQEAHGNKWPQVSLRANDTYNMHFCGGS 60  
Qy 79 LLTSRWVITAAHC-----FKNLNKPFLFSVLLGAWLGNFGSRQSKYGVAMVE 128  
||: ||:||||| : | | | | :

Db 61 LIHPQWVLTAAHCVPDVPADPNKVRVOLRKQYLY-----YHDHLMTVSQII 107  
Qy 129 HP-VYSWKEGACADIALVRLERSIQSERVLPICLPDASIHLPNTHCWISWGSIDQGV 187  
|| | :|| ||||| : | : ||||| | | ||:||||| :  
Db 108 HPDYIIVQDG--ADIALKLITNPVNI SDYVHPVLPPLPASETFPSGTLCTWGTGNGNIDNGV 165  
Qy 188 PLHPOTLOKLVPIIDSEVCSHLYWRGAGQGP-----ITEDMLCAGYLEGERDACLGDSG 243  
|| | :|| ||||| : | : ||||| | | :|||  
Db 166 NLPPFPPLKEVQVPIIENHLCDLKYLKGLTGNVHVIRDDMLCAGN-EG-HDSCQGDG 223  
Qy 244 GGLMCQVDGAWLLAGIISWEGGCAERNRPGVYISLSAHSRWSWEKIV 289  
|||:|:| | ||:||||||| ||||| : : | :  
Db 224 GGLVCKVEDTWLQAGVYVSWEGGCAQPNRPGIYTRVYIYLDWIHRYV 269

## RESULT 15

US-09-016-366A-23  
; Sequence 23, Application US/09016366A  
; Patent No. 5955431  
; GENERAL INFORMATION:  
; APPLICANT: Stevens, Richard L.  
; APPLICANT: Huang, Chifu  
; TITLE OF INVENTION: MAST CELL PROTEASE PEPTIDE  
; TITLE OF INVENTION: INHIBITORS  
; NUMBER OF SEQUENCES: 65  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.  
; STREET: 600 Atlantic Avenue  
; CITY: Boston  
; STATE: MA  
; COUNTRY: U.S.A.  
; ZIP: 02210-2211  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FASTSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/016,366A  
; FILING DATE: January 30, 1998  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/037,090  
; FILING DATE: 05-FEB-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Plumer, Elizabeth R.  
; REGISTRATION NUMBER: 36,637  
; REFERENCE/DOCKET NUMBER: B0801/7093  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617-720-3500  
; TELEFAX: 617-720-2441  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 23:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 267 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: No. 5955431e  
US-09-016-366A-23

Query Match 29.9%; Score 514.5; DB 2; Length 267;  
Best Local Similarity 38.8%; Pred. No. 2.5e-43;  
Matches 107; Conservative 51; Mismatches 83; Indels 35; Gaps 8;  
Qy 32 AARIPVPPACKPQOLNRVVGEDSTSEWPIVSIQ---KNGTHHCAGSLLTSRWVITA 88  
|| | :|| | | :||| :||| ||| : | |||  
Db 5 ASRAYAAPAPQALQRYGVIGGQEA RSKWPQVSLVRDRYWMHFCGSLIHPQWVILTA 64  
|| | :|| | | :||| :||| ||| : | |||  
Qy 89 AHC-----FKD-----NLNKPFLFSVLLGAWLGNFGSRQSKYGVAMVEPHP-VYSWKEG 137  
||| ||| ||| :||| :||| :||| :||| :|||  
Db 65 AHCVPDVKDLAALRLVQLREQHLY-----YQDQLLPVSIIVHPQFYTAQIG 111  
||: ||:||||| : | | | | :

Qy 138 ACADIALVRLERSIQFSERVLPICLPDASHLPENTHCWISGWSIODGYPLPHPTLQK 197  
Db 112 --ADIALLELEEPVKVSSHVHTVTLPPASETFPPGMPWCWVTGWGDVNDNERLPPFPPLKQ 169  
Qy 198 LKVPIDSEVCSHLYWRGAGQGP----ITEDMLCAGYLEGERDACLGDGSGPLMCQVDGA 253  
Db 170 VKVPIMENHICDAKYHLGAYTGDDVRIVRDDMLCAG--NTRRDSQGDGSGGPLVCKVNGT 227  
Qy 254 WLLAGIISWEGGCAERNRPGVYISLSAHRSWVEKIV 289  
Db 228 WLQAGVYSWEGGCAQPNRPGIYTRVYIYLDWIHHYV 263

Search completed: August 29, 2002, 10:24:28  
Job time: 722 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2002, 10:13:56 ; Search time 17.52 Seconds  
(without alignments)  
1738.603 Million cell updates/sec

Title: US-09-903-925a-263  
Perfect score: 1720  
Sequence: 1 MVVSGAPPALGGGCLGTFTS.....AQQGGALRAPSGSGAAARS 317

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_71.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	607.5	35.3	343	1 A57014	proctasin (EC 3.4.21)
2	554	32.2	276	2 A38654	mast cell proteina
3	546	31.7	275	2 A32410	tryptase (EC 3.4.21)
4	538.5	31.3	274	2 JC4171	tryptase (EC 3.4.21)
5	536	31.2	237	2 S68702	tryptase (EC 3.4.21)
6	532	30.9	274	2 A45754	tryptase (EC 3.4.21)
7	530.5	30.8	275	2 B35863	tryptase (EC 3.4.21)
8	529.5	30.8	275	2 A35863	tryptase (EC 3.4.21)
9	522.5	30.4	275	2 C35863	tryptase (EC 3.4.21)
10	522	30.3	270	2 S56160	mast cell tryptase
11	518	30.1	273	2 A47246	tryptase (EC 3.4.21)
12	481	28.0	638	1 KQHUP	plasma kallikrein
13	475.5	27.6	625	1 KFHUI	coagulation factor
14	473	27.5	638	1 KQMPL	plasma kallikrein
15	469.5	27.3	638	1 KQTPPL	plasma kallikrein
16	460.5	26.8	812	1 PLBO	plasmin (EC 3.4.21)
17	457.5	26.6	1034	1 A53663	enteropeptidase (E
18	452	26.3	417	1 S00845	hepsin (EC 3.4.21)
19	450	26.2	416	1 S33777	hepsin (EC 3.4.21)
20	450	26.2	810	2 B30848	plasmin (EC 3.4.21)
21	446	25.9	455	2 A61545	plasmin (EC 3.4.21)
22	446	25.9	790	1 PLPG	plasmin (EC 3.4.21)
23	446	25.9	810	1 PLHU	plasmin (EC 3.4.21)
24	445.5	25.9	1524	2 T30337	polypeptide - Afri
25	445	25.9	786	1 A47347	serine proteinase
26	444	25.8	1019	1 A56318	enteropeptidase (E
27	443.5	25.8	1035	1 A43090	enteropeptidase (E
28	442	25.7	1420	2 A32869	apolipoprotein(a)
29	441	25.6	460	2 B61545	plasmin (EC 3.4.21)

30	439.5	25.6	812	1 PLMS	plasmin (EC 3.4.21)
31	437.5	25.4	367	2 JE0104	testicular serine
32	437.5	25.4	437	2 S18407	acrosin (EC 3.4.21)
33	437.5	25.4	4548	1 S00657	apoptotrypsin(a) (EC
34	437	25.4	263	1 KQRTB	chymotrypsin (EC 3
35	437	25.4	421	1 S11674	acrosin (EC 3.4.21)
36	434	25.2	265	2 T15451	hypothetical prote
37	433.5	25.2	271	2 A25528	pancreatic elastas
38	432.5	25.1	436	2 JX0172	testicular serine
39	430.5	25.0	366	2 JE0105	testicular serine
40	429.5	25.0	431	2 S47538	acrosin (EC 3.4.21)
41	428	24.9	263	2 A21195	chymotrypsin (EC 3
42	425	24.7	810	2 I46260	plasmin (EC 3.4.21)
43	424.5	24.7	270	2 B29934	pancreatic elastas
44	422	24.5	263	2 A31299	chymotrypsin (EC 3
45	422	24.5	415	1 A34170	acrosin (EC 3.4.21)

ALIGNMENTS

RESULT 1  
A57014  
proctasin (EC 3.4.21.-) precursor - human  
C:Species: Homo sapiens (man)  
C:Date: 24-May-1996 #sequence\_revision 24-May-1996 #text\_change 18-Jun-1999  
C:Accession: A57014; A54866  
R;Yu, J.X.; Chao, L.; Chao, J.  
J. Biol. Chem. 270, 13483-13489, 1995  
A:Title: Molecular cloning, tissue-specific expression, and cellular localization of  
A:Reference number: A57014; MUID:95286644  
A:Accession: A57014  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-343 <RES>  
A:Cross-references: GB:I41351; NID:g862304; PIDN:AAC41759.1; PID:g862305  
A:Experimental source: prostate  
A:Note: parts of this sequence were determined by protein sequencing  
R;Yu, J.X.; Chao, L.; Chao, J.  
J. Biol. Chem. 269, 18843-18848, 1994  
A:Title: Proctasin is a novel human serine proteinase from seminal fluid. Purificatio  
A:Reference number: A54866; MUID:94308140  
A:Accession: A54866  
A:Molecule type: protein  
A:Residues: 45-64 <YDA>  
C:Genetics:  
A:Gene: GDB:PRSS8  
A:Cross-references: GDB:676446; OMIM:600823  
A:Map position: 16p11.2-16p11.2  
C:Superfamily: proctasin; trypsin homology  
C:Keywords: glycoprotein; hydrolase; serine proteinase; transmembrane protein  
F:1-32/Domain: signal sequence #status predicted <SIG>  
F:33-44,45-343/Product: proctasin #status predicted <MAT>  
F:33-44/Domain: proctasin light chain #status predicted <CHL>  
F:45-343/Domain: proctasin heavy chain #status predicted <CHH>  
F:45-281/Domain: trypsin homology <TRY>  
F:323-341/Domain: transmembrane #status predicted <TMM1>  
F:37-154,70-86,168-244,201-223,234-262/Disulfide bonds: #status predicted  
F:85,134,238/Active site: His, Asp, Ser #status predicted  
F:159/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 35.3%; Score 607.5; DB 1; Length 343;  
Best Local Similarity 44.2%; Pred. No. 6.3e-44;  
Matches 129; Conservative 42; Mismatches 108; Indels 13; Gaps 7;

QY	10	LGSGCLGTFSTLL---LLASTAILNAARIPVPPACGKQQLNRVVGSGEDSTDSEMPWIVS	66
Db	7	LGPGQLGAVAILLYGLLRSGTGAGRAP---CGVAPQA-RITGSSAVAGQMPWQVS	61
QY	67	IQKNGTHHCAGSLTSRWVITAAHCFKDNLNKPYLFVYLLGAWQLGNPGRSQKVGVAWV	126
Db	62	ITYEGVHVGGSLSYSEQWVLSAAHCFPSEHKE-AYEVKLGHAHQLDYSSEDAKYSTLKDI	120

[illegible]

RESULT 2

A38654

muscle cell proteinase 6 (EC 3.4.21.-) precursor - mouse

C:Species: Mus musculus (house mouse)

C:Date: 21-Feb-1992 #sequence-revision 17-Feb-1994 #text\_change 22-Jun-1999

C:Accession: A38654; B38654; D35646; I59478

R:Reynolds, D.S.; Gurley, D.S.; Austen, K.F.; Serafin, W.E.

J. Biol. Chem. 266, 3847-3853, 1991

A:Title: Cloning of the cDNA and gene of mouse mast cell protease-6. Transcription by

A:Reference number: A38654; MUID:91139682

A:Accession: A38654

A:Molecule type: DNA

A:Residues: 1-276 <REY>

A:Cross-references: GB:M57625; NID:Q200506; PIDN:AAA39987.1; PID:Q200507

A:Note: the authors translated the codon CCG for residue 24 as Ala, GAG for residue 37 as Gly, GAG for residue 148 as Gly, GAG for residue 168 as Gly, and GAA for 185 as Gly

A:Accession: B38654

A:Molecule type: mRNA

A:Residues: 1-276 <RE2>

A:Cross-references: GB:M57626; NID:Q200508; PIDN:AAA39988.1; PID:Q200509

R:Reynolds, D.S.; Stevens, R.L.; Lane, W.S.; Carr, M.H.; Austen, K.F.; Serafin, W.E.

Proc. Natl. Acad. Sci. U.S.A. 87, 3230-3234, 1990

A:Title: Different mouse mast cell populations express various combinations of at least

A:Reference number: A35646; MUID:9022202

A:Accession: D35646

A:Molecule type: protein

A:Residues: 32-54 <RE3>

R:Huang, R.; Abrink, M.; Gobl, A.E.; Nilsson, G.; Aveskogh, M.; Larsson, L.G.; Nilsson, S.

Scand. J. Immunol. 38, 359-367, 1993

A:Title: Expression of a mast cell tryptase in the human monocytic cell lines U-937 and

A:Reference number: I59478; MUID:94023807

A:Accession: I59478

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-276 <RES>

A:Cross-references: GB:L31853; NID:Q473480; PIDN:AAA39725.1; PID:Q473481

C:Genetics:

A:Gene: MNCP-6

A:Introns: 24/1; 79/2; 168/1; 222/3

C:Superfamily: trypsin; trypsin homology

C:Keywords: hydrolase; serine proteinase; zymogen

F:1-21/Domain: signal sequence #status predicted <SIG>

F:22-31/Domain: activation peptide #status predicted <ACT>

F:32-276/Product: mast cell proteinase 6 #status experimental <MAT>

F:32-268/Domain: trypsin homology <TRY>

F:75; 122, 225/Active site: His, Asp, Ser #status predicted

```

Qy 135 KEGACADIALVRLERSIQFSERVLPICLDPASTHLPNTHCWSGSGTQDGVPLPHPQT 194
      |||||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 117 AEGG-ADVALLLEVPNVNSTHPISLPPASSTFPFGTSCWVTGWGDINDDEPLPPYP 175
      |||||::||::||::||::||::||::||::||::||::||::||::||::||::||
Qy 195 LOKLVPIIDSEVCSHLYNRGAQG---PITED--MLCAGYLEGERDACILGGGLMCOV 250
      |||||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 176 LKQVKVIPENSLCDRKYHTGLTGDDFPFIVHDGMLCAG--NTRRDSOQDSSGGLPVCKY 233
      |||||::||::||::||::||::||::||::||::||::||::||::||::||::||
Qy 251 DGAWLLAGIISWGEGAERNRPQVYISLSAHRSSWEKIV 289
      |||||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 234 KGTVLOAGVYSWGEGCAOPNKPKPIYTRVYLIDWIHRYV 272
      |||||::||::||::||::||::||::||::||::||::||::||::||::||::||

```

```

RESULT      3
A32410
  tryptase (EC 3.4.21.59) precursor - dog
C:Species: Canis lupus familiaris (dog)
C:Date: 12-Oct-1989 #sequence_revision 12-Oct-1989 #text_change 22-Jun-1999
C:Accession: A32410
R:Vanderslice, P.; Craik, C.S.; Nadel, J.A.; Caughey, G.H.
Biochemistry 28, 4148-4155, 1989
A:Title: Molecular cloning of dog mast cell tryptase and a related protease
A:Reference number: A32410; MUID:89352460
A:Accession: A32410
A:Molecule type: mRNA
A:Residues: 1-275 <VAN>
A:Cross-references: GB:M24664; NID:q163982; PIDN:AAA30854.1; PID:g163983;
C:Superfamily: trypsin; trypsin homology
C:Keywords: hydrolase; serine protease; zymogen
F:1-20/Domain: signal sequence status predicted <SIG>
F:22-30/Domain: activation peptide status predicted <ACT>
F:31-275/Product: tryptase status predicted <MAP>
F:31-267/Domain: trypsin homology <TRY>
F:74,121,224/Active site: His, Asp, Ser #status predicted

  Query Match      31.7%; Score 546; DB 2; Length 275;
  Best Local Similarity 40.9%; Pred. No. 8e-39;
  Matches 112; Conservative 46; Mismatches 102; Indels 14; Gaps

QY 23 LLASTAILNARIPVPPACGKQFQQLNRVVGGEDSTDSEKWPWIVSTOKNG---THHCAGSL 79
Db 5 LVLVALALLGSL-VPVSPAPQALQRVIGVGGREAPGSKWPQVQVSLRLKQGYWRHICGSL 63
QY 80 LTSRWITTAACPKDNLNPKYFLSVLLGAQWLGNPGSRSQKGVGVAWPEHPVYSWKEGAC 139
Db 64 IHPQWVLTAAHCVPNVCPEEIRVOLRQHL---YYQDHLPLVNRVIVHNPVYTPENG- 119
QY 140 ADIALVRLERSIOFSRVLPICLPDASIHLPNTTHCWISGMSIGDGVPLPHPTQLOKLK 199
Db 120 ADIALLEEDPVNVAHQVPTLPPALQFTPTGTCWWTGWDVHSQTPLEPPPLKQVK 179
QY 200 VPIIDSEVCSHLYWRGAGOGP-----ITEDMLCAGYLEGERDACLDGSGGLPMQVDGAWL 255
Db 180 VPIVENSMDQVQVHLGLSTGDGVRIVRVREDMLCAG--NSKSDSCQDGGPLVCRVRGVWL 237
QY 256 LAGIIISWEGGCAERNRPPGVYIISLSAHRSWEKIV 289
Db 238 QAGVWSGEGCAQPNRPGIYTRVAYLDWIHQYV 271

RESULT      4
JC4171
  tryptase (EC 3.4.21.59) precursor - rat
N:Alternate names: mast cell tryptase
C:Species: Rattus norvegicus (Norway rat)
C:Date: 27-Aug-1995 #sequence_revision 27-Oct-1995 #text_change 20-Jun-2000
C:Accession: JC4171
R:Ide, H.; Itoh, H.; Tomita, M.; Murakumo, Y.; Kobayashi, T.; Maruyama, H.;
J. Biochem. 118, 210-215, 1995
A:Title: cDNA sequencing and expression of rat mast cell tryptase.
A:Reference number: JC4171; MUID:96015171

```

Qy	116	SRSQVGVAVWEHPH - YYSWKEGACADIALVRLERSIQFSERVLPICLPDASIHLPNPTH	174
		: : : :                 : : :	
Db	59	YQDQLLPISRIIPHPNCYSVKNG - ADIALLELDKLVNISSWHVQVTLPPSESTFPPEGTQ	116
Qy	175	CWISGWSIGDGVPLPHPOFLQKLKVPDIIDSEVCSHLYWRGACQGP - ----ITEDMLCAGY	230
		: : :                 : : :	
Db	117	CWITGWNVDNGRRLPPFPFLQKVQVPEVNSVCDRKYHSGLSTGDNVPIVREDMLCAG -	175
Qy	231	LEGEDRACLDSGGPLMCQVDGAWLLAGIISWGEGCAERNRPGVYISLSAHSRWSVEKIV -	289
		: :             : :             : :             : :	
Db	176	-DSGRNFCQDSSGGLPCKYNGTWLQGVVSWGEGCAKPNRPGLYTRVYSYLDWIHQYVP	234
Qy	290	QG 291	
Db	235	QG 236	

```

RESULT      6
A45754
trypsinase (EC 3.4.21.59) alpha precursor - human
C/Species: Homo sapiens (man)
C/Date: 03-Jun-1993 #sequence_revision 03-Jun-1993 #text_change 08-Sep-1997
C/Accession: A45754; B37193
R/Miller, J.S.; Westin, E.H.; Schwartz, L.B.
J. Clin. Invest. 84, 1188-1195, 1989
A/Title: Cloning and characterization of complementary DNA for human trypsinase.
A/Reference number: A45754; MUID:90009311
A/Accession: A45754
A/Molecule type: mRNA
A/Residues: 1-274 <MI1>
A/Cross-references: GB:M30038
R/Miller, J.S.; Moxley, G.; Schwartz, L.B.
J. Clin. Invest. 86, 864-870, 1990
A/Title: Cloning and characterization of a second complementary DNA for human trypsinase.
A/Reference number: A37193; MUID:90369005
A/Accession: B37193
A/Molecule type: mRNA
A/Residues: 1-274 <MI2>
A/Cross-references: GB:M30038

```

	C;Superfamily: trypsin;	trypsin homology	
	C;Keywords: hydrolase;	serine proteinase; zymogen	
F;1-21/Domain:	signal sequence	#status predicted <SIG>	
F;22-30/Domain:	activation peptide	#status predicted <ACT>	
F;31-274/Product:	trypsin I	#status predicted <MAT>	
F;31-286/Domain:	trypsin homology	<TRY>	
F;74,120,223/Active site:	His, Asp, Ser	#status predicted	
Query Match	30.9%;	Score 532;	DB 2; Length 274;
Best Local Similarity	40.1%;	Pred. No. 1.2e-37;	
Matches 113;	Conservative 50;	Mismatches 93;	Indels 26; Gaps
QY	20 SLLLSATLNAARI	VPVPCGKPOQLNRVVGEDSTDSEWPWISIQ---	KNGTHHCA 76
	:	:                     :	:
Dd	3 SLLLTALPV-	--ASRAVAAPVALQAGIVGGQAPRSKPWQSLRVDRYIWNHFCG 60	
QY	77 GSLLTSRWVTIAAHC	RFDNLNKKPYLFGLLGAWQLNGPS----	RSQKVGVAVVEPH-V 131
	:                     :	:	:
Dd	61 GSLTHPQWLTAAC	LCPDVKDLATLV-----	NSGTHLYYQQDLLPVSRIMVHPQF 112
QY	132 YSWKEGACADIAL	VRLERSIOFSERVLPICLPDASTHLPPNTHCWTSGWGSIQDGVP LPH 191	
	:                         :	:	:
Dd	113 YIIOTG--ADIALE	EPEPNYSRRVHTMCLPASETFPGMCWTVTGWDVDNDPLPP 170	

Qy	192	PQTQLKLVPIIDSEVCSHLYWRGAGQP----	ITDMLCAGYLEGERDACLDGSGPLM	247
Db	171	PFPLQKVQVPMENHHICDAKYHLGATGDDVRI	IRDMLCAG--NSORDCKGDSGPLV	228
Qy	248	QVDGAWLLAGIISWGEGCAERNRPGVYISLSA	HRSVKIV	289
Db	229	CKVNGTWLQAGVSWDEGCAQPNRPGIYTRVY	YLDWIHHV	270









QY 49 RVVGGEDSTDSEWPVIYSIOKNG--THHCAGSLTTSRWVITAAHCFKDNLNKPYLFSVL 105  
 Db 387 RIVGGTASVGEPMQVTLHTTSTORHLCGGSLIGNQWILTAHCFY-GVESPKILRVY 445  
 QY 106 LGAWOLNPGSRKOVGVAVVEHPVTSWKEGACADIALVRLERSIOFSERVLPICLPDA 165  
 Db 446 SGILNQSEIKEDTSFFGVQEIHHQYKMAESG-YDIALKLLETVNYTDSQRPICLP 504  
 QY 166 SIHLPPNTHCWISWG--STODGVPLPHQOTLOKLVPIIDSEVCSHLXWRGAGQGPFITE 223  
 Db 505 GDRNVIYTDQVWTGWYKRLDKT---QNTLQAKIPLVTNECQKRY-RG---HKITH 556  
 QY 224 DMLCAGYLEGERDACLDGSGPLMCOVDGAWLLAGIISWEGCAERNRPGVYISLSHRS 283  
 Db 557 KMCAGYREGGDKACKDGGSLGCKHNEYHLVIGTSGEGCAQRPGRPGVYVVEYVD 616  
 QY 284 WVEKIVGV 292  
 Db 617 WLEKTOAV 625

RESULT 14  
 KQMSPL  
 A:plasma kallikrein (EC 3.4.21.34) precursor - mouse  
 C:Species: Mus musculus (house mouse)  
 C>Date: 30-Sep-1992 #sequence\_revision 30-Sep-1992 #text\_change 18-Jun-1999  
 C:Accession: A36557  
 R:Seidah, N.G.; Sawyer, N.; Hamelin, J.; Mion, P.; Beaubien, G.; Brachpapa, L.; Rochemont  
 DNA Cell Biol. 9, 737-748, 1990  
 A:Title: Mouse plasma kallikrein: cDNA structure, enzyme characterization, and comparison  
 A:Reference number: A36557; MUID:91090844  
 A:Accession: A36557  
 A:Molecule type: mRNA  
 A:Residues: 1-638 <SEI>  
 A:Cross-references: GB:M58588; NID:g200358; PIDN:AAA63393.1; PID:g200359  
 A:Note: part of this sequence, including the amino ends of both the heavy and light chain  
 C:Comment: This protein, synthesized in the liver, circulates as a noncovalent complex w  
 are linked by one or more disulfide bonds.  
 C:Superfamily: coagulation factor XI; trypsin homology  
 C:Keywords: blood coagulation; duplication; fibrinolysis; glycoprotein; hydrolase; infla  
 F:1-19/Domain: signal sequence #status predicted <SIG>  
 F:20-30/Product: plasma kallikrein heavy chain #status experimental <HCH>  
 F:20-109/Domain: apple repeat <AP1>  
 F:110-199/Domain: apple repeat <AP2>  
 F:200-289/Domain: apple repeat <AP3>  
 F:291-380/Domain: apple repeat <AP4>  
 F:391-638/Product: plasma kallikrein light chain #status experimental <LCH>  
 F:391-621/Domain: trypsin homology <TRY>  
 F:21-104,47-77,51-57,111-194,137-166,141-147,201-284,227-256,231-237,292-375,318-347,322  
 F:127,215,308,396,494/Binding site: carbohydrate (Asn) (covalent) #status predicted  
 F:434,483,578/Active site: His, Asp, Ser #status predicted

Query Match 27.5%; Score 473; DB 1; Length 638;  
 Best Local Similarity 35.3%; Pred. No. 3.3e-32;  
 Matches 106; Conservative 54; Mismatches 108; Indels 32; Gaps 10;

QY 5 GAPPALGGCGLGTFTSLLLASTAILNAARIPVPACGKQQLN-RVVGGEDSTDSEWP 63  
 Db 355 GSPRTIYGMQ-----SSGVSRLCKLVDPDC--TTKINARIVGGTNASLGEPW 404  
 QY 64 IVSTQ----KNGTHHCAGSLTTSRWVITAAHCFKDNLNKPYLFSVLGAWQLG-----NPG 115  
 Db 405 QVSLQKLVSTHLCGGSLIGNQWILTAHCF-DGIPYDWRVYGGILSLEITKTPS 463  
 QY 116 SRSKOVGVAVVEHPVTSWKEGACADIALVRLERSIOFSERVLPICLPDASIHLPNTHC 175  
 Db 464 SRIKELII-----HQEYKVSSEGN-YDIALKLQPLNTYEFQKPICLPSKADNTIYTC 517  
 QY 176 WISGWSIQDGVPLPHQOTLOKLVPIIDSEVCSHLXWRGAGQGPFITDMLCAGYLEGER 235

Db 518 WVTGWGYTKEOGTQN--ILQKATIPLPVNECQKKY----RDVYINKQMICAGYKEGGT 571  
 QY 236 DACLDGSGGGLMCOVDGAWLLAGIISWEGCAERNRPGVYISLSHRSWKEIVQGVQLR 295  
 Db 572 DACKGDSGGPLVCKHSGRWQLVIGTSGEGCRKDQPGVYTKVSEYMDWILEKTQSSDVR 631

RESULT 15  
 KQRTPL  
 A:plasma kallikrein (EC 3.4.21.34) precursor - rat  
 N:Alternate names: Fletcher factor; kininogenin; serum kallikrein  
 C:Species: Rattus norvegicus (Norway rat)  
 C>Date: 30-Sep-1992 #sequence\_revision 30-Sep-1992 #text\_change 18-Jun-1999  
 C:Accession: A39180; A33320; S06851; I53041; S06852  
 R:Beaubien, G.; Rosinski-Chupin, I.; Mattel, M.G.; Mbikay, M.; Chretien, M.; Seidah,  
 Biochemistry 30, 1628-1635, 1991  
 A:Title: Gene structure and chromosomal localization of plasma kallikrein.  
 A:Reference number: A39180; MUID:91129236  
 A:Accession: A39180  
 A:Molecule type: DNA  
 A:Residues: 1-638 <BEA>  
 A:Cross-references: GB:J05315  
 A:Note: the authors translated the codon GAG for residue 81 as Gln  
 R:Seidah, N.G.; Ladenheim, R.; Mbikay, M.; Hamelin, J.; Lutfalla, G.; Rougeon, F.; La  
 DNA 8, 563-574, 1989  
 A:Title: The cDNA structure of rat plasma kallikrein.  
 A:Reference number: A33320; MUID:90091743  
 A:Accession: A33320  
 A:Status: not compared with conceptual translation  
 A:Molecule type: mRNA  
 A:Residues: 1-638 <SEI>  
 A:Cross-references: GB:M30282; NID:g205010; PIDN:AAA41463.1; PID:g205011  
 A:Note: part of this sequence, including the amino ends of both the heavy and light c  
 R:Paquin, J.; Benjannet, S.; Sawyer, N.; Lazure, C.; Chretien, M.; Seidah, N.G.  
 Biochim. Biophys. Acta 999, 103-110, 1989  
 A:Title: Rat plasma kallikrein: purification, NH(2)-terminal sequencing and developme  
 A:Reference number: S06851; MUID:90089457  
 A:Accession: S06851  
 A:Molecule type: protein  
 A:Residues: 20-45;391-413 <PAQ>  
 R:Seidah, N.G.; Ladenheim, R.; Mbikay, M.; Hamelin, J.; Lutfalla, G.; Rougeon, R.; La  
 DNA Cell Biol. 8, 563-574, 1989  
 A:Title: The cDNA structure of rat plasma kallikrein.  
 A:Reference number: I53041  
 A:Accession: I53041  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-638 <RES>  
 A:Cross-references: GB:M58590; NID:g206721; PIDN:AAA42069.1; PID:g206722  
 C:Comment: This protein, synthesized in the liver, circulates as a noncovalent comple  
 are linked by one or more disulfide bonds.  
 C:Genetics:  
 A:Gene: PK  
 C:Superfamily: coagulation factor XI; trypsin homology  
 C:Keywords: blood coagulation; duplication; fibrinolysis; glycoprotein; hydrolase; in  
 F:1-19/Domain: signal sequence #status predicted <SIG>  
 F:20-390/Product: plasma kallikrein heavy chain #status experimental <MAT1>  
 F:20-109/Domain: apple repeat <AP1>  
 F:110-199/Domain: apple repeat <AP2>  
 F:200-289/Domain: apple repeat <AP3>  
 F:291-380/Domain: apple repeat <AP4>  
 F:391-638/Product: plasma kallikrein light chain #status experimental <MAT2>  
 F:391-621/Domain: trypsin homology <TRY>  
 F:21-104,47-77,51-57,111-194,137-166,141-147,201-284,227-256,231-237,292-375,318-347,  
 F:127,215,308,453,459,494/Binding site: carbohydrate (Asn) (covalent) #status predict  
 F:396/Binding site: carbohydrate (Asn) (covalent) #status experimental  
 F:434,483,578/Active site: His, Asp, Ser #status predicted

Query Match 27.3%; Score 469.5; DB 1; Length 638;  
 Best Local Similarity 37.6%; Pred. No. 6.5e-32;  
 Matches 94; Conservative 50; Mismatches 95; Indels 11; Gaps 5;

```

QY 49 RVVGEDSTDSEWPVIVSIQ---KNGTHHCAGSLTSSRVITAAHCFKMDNLNKPFLFSVL 105
Db 390 RIVGTNSSLGSEWPQVSLQVKLVSONHMCSSIIIGRQWILTAHCF-DGIPYDPVWRIY 448
QY 106 LGAWOLGNPGSRSQKVGVAWVEPHVPVYSWKEGACADIALVRLERSIQFSERVLPICLPDA 165
Db 449 GGILNLSEITNKTPESSIKELIIHQKYKMEGS-YDIALIKLQTPLNYTEFQKPICLPFSK 507
QY 166 STHLPNTHCWTSGWSTQDGVPLPHPOTLOKLVPIIDSEVCSHLYWRGAGQGPTEDM 225
Db 508 ADTNTIYNCWVTGWYTKERGETQN--ILOKATIPLVNEECOKKY----RDYVITKQM 561
QY 226 LCAGYLEGERDACLDGSGPLMCQVDGAWLLAGIISWEGECAERNRPGVYISLSAHRSWV 285
Db 562 ICAGYKEGGIDACKDGGPLVCKRHSGRWQLVGITSWEGEGCARKEQPGVYTKVAEYIDWI 621
QY 286 EKIVQGVQLR 295
Db 622 LEKIQSSKER 631

```

Search completed: August 29, 2002, 10:24:52  
Job time: 656 sec



```
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
KW PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolyase; Serine protease; Signal.
FT SIGNAL 1 32 POTENTIAL.
FT CHAIN 33 317 BRAIN-SPECIFIC SERINE PROTEASE 4.
FT ACT_SITE 90 90 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 141 141 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 242 242 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 75 91 BY SIMILARITY.
FT DISULFID 175 248 BY SIMILARITY.
FT DISULFID 208 227 BY SIMILARITY.
FT DISULFID 238 266 BY SIMILARITY.
FT CARBOHYD 70 70 N-LINKED (GLCNAC... ) (POTENTIAL).
FT CONFLICT 47 47 L -> M (IN REF. 3).
SQ SEQUENCE 317 AA; 33731 MW; E2A123BC86E79935 CRC64;

Query Match 100.0%; Score 1720; DB 1; Length 317;
Best Local Similarity 100.0%; Pred. No. 2.6e-141;
Matches 317; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVVSGAPPALGGCLGFTSLLLAATAILNAARIPVPACGKPKQQLNRVVGGEDSTDSE 60
DB 1 MVVSGAPPALGGCLGFTSLLLAATAILNAARIPVPACGKPKQQLNRVVGGEDSTDSE 60
QY 61 WPWIVSIQKNGTHRCAGSLTSTRWVITAAHCFKDNLNKPYLFSVLLGAWQLGNPGRSQK 120
DB 61 WPWIVSIQKNGTHRCAGSLTSTRWVITAAHCFKDNLNKPYLFSVLLGAWQLGNPGRSQK 120
QY 121 VGVAVPEHPVYSKKEGACADIALVRLERSIQFSERVLPICLPDASHLPPNTHCWISGW 180
DB 121 VGVAVPEHPVYSKKEGACADIALVRLERSIQFSERVLPICLPDASHLPPNTHCWISGW 180
QY 181 GSIQDGVPLPQTLQKLVPIIDSEVCSHLYWRGAGGPTEDMLCAGYLEGERDACL 240
DB 181 GSIQDGVPLPQTLQKLVPIIDSEVCSHLYWRGAGGPTEDMLCAGYLEGERDACL 240
QY 241 DSGGPLMCQVDGAWLLAGIISWEGGCAERNRPGVYISLSAHRSWVEKIVQVQLRGAQG 300
DB 241 DSGGPLMCQVDGAWLLAGIISWEGGCAERNRPGVYISLSAHRSWVEKIVQVQLRGAQG 300
QY 301 GGALRAPSGSGAAARS 317
DB 301 GGALRAPSGSGAAARS 317

RESULT 2
BSS4_MOUSE
ID BSS4_MOUSE STANDARD; PRT; 306 AA.
AC Q9R10;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Brain-specific serine protease 4 precursor (EC 3.4.21.-) (BSSP-4).
GN PRSS26 OR BSSP4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Mitsui S., Okui A., Kominami K., Yamaguchi N.;
RT "Cloning and characterization of a novel serine protease, mBSSP-4.";
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
```

```
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB010778; BAB20262.1; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolyase; Serine protease; Signal.
FT SIGNAL 1 32 POTENTIAL.
FT CHAIN 33 306 BRAIN-SPECIFIC SERINE PROTEASE 4.
FT ACT_SITE 90 90 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 141 141 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 242 242 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 75 91 BY SIMILARITY.
FT DISULFID 175 248 BY SIMILARITY.
FT DISULFID 208 227 BY SIMILARITY.
FT DISULFID 238 266 BY SIMILARITY.
FT CARBOHYD 70 70 N-LINKED (GLCNAC... ) (POTENTIAL).
SQ SEQUENCE 306 AA; 33262 MW; FBBF03C0C285E7E8 CRC64;

Query Match 73.9%; Score 1270.5; DB 1; Length 306;
Best Local Similarity 76.8%; Pred. No. 1.5e-102;
Matches 232; Conservative 25; Mismatches 44; Indels 1; Gaps 1;

QY 1 MVVSGAPPALGGCLGFTSLLLAATAILNAARIPVPACGKPKQQLNRVVGGEDSTDSE 60
DB 1 MMISRRPPALGGDQFSILILLVLTSTAPISAATIRVSPDCGKPKQQLNRIVGGEDSMDAQ 60
QY 61 WPWIVSIQKNGTHRCAGSLTSTRWVITAAHCFKDNLNKPYLFSVLLGAWQLGNPGRSQK 120
DB 61 WPWIVSIQKNGTHRCAGSLTSTRWVITAAHCFKDNLNKPYLFSVLLGAWQLGNPGRSQK 120
QY 121 VGVAVPEHPVYSKKEGACADIALVRLERSIQFSERVLPICLPDASHLPPNTHCWISGW 180
DB 121 VGVAVPEHPVYSKKEGACADIALVRLERSIQFSERVLPICLPDASHLPPNTHCWISGW 180
QY 181 GSIQDGVPLPQTLQKLVPIIDSEVCSHLYWRGAGGPTEDMLCAGYLEGERDACL 240
DB 181 GSIQDGVPLPQTLQKLVPIIDSEVCSHLYWRGAGGPTEDMLCAGYLEGERDACL 240
QY 241 DSGGPLMCQVDGAWLLAGIISWEGGCAERNRPGVYISLSAHRSWVEKIVQVQLRGAQG 300
DB 241 DSGGPLMCQVDGAWLLAGIISWEGGCAERNRPGVYISLSAHRSWVEKIVQVQLRGAQG 300
QY 301 GG 302
DB 300 SG 301

RESULT 3
MPN_HUMAN
ID MPN_HUMAN STANDARD; PRT; 290 AA.
AC Q9BQR3;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Marapsin precursor (EC 3.4.21.-).
GN MPN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Fortunato M., Dando P.M., Rawlings N.D., Barrett A.J.;
RT "Cloning, sequencing and expression of marapsin, a human serine
RT proteinase.";
```

RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
 CC TRYPSIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AJ306593; CAC35467.1; -  
 CC MEROPS: S01.074; -  
 CC InterPro: IPR001314; Chymotrypsin.  
 CC InterPro: IPR001254; Trypsin.  
 CC Pfam: PF00089; trypsin; 1.  
 CC PRINTS: PR00722; CHYMOTRYPSIN.  
 CC SMART: SM00020; Tryp\_SPC; 1.  
 CC PROSITE: PS00240; TRYPSIN\_DOM; 1.  
 CC PROSITE: PS00134; TRYPSIN\_HIS; 1.  
 CC PROSITE: PS00135; TRYPSIN\_SER; 1.  
 CC Hydrolase; Serine protease; zymogen; Signal; Glycoprotein.  
 KW SIGNAL 1 22 POTENTIAL.  
 FT PROPEP 23 34 MARAPSN.  
 FT CHAIN 35 290  
 FT DOMAIN 35 277  
 FT ACT\_SITE 75 75 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 124 124 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 229 229 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT DISULFID 60 76 BY SIMILARITY.  
 FT DISULFID 158 235 BY SIMILARITY.  
 FT DISULFID 191 214 BY SIMILARITY.  
 FT DISULFID 225 253 BY SIMILARITY.  
 FT CARBOHYD 55 55 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 79 79 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 290 AA; 31940 MW; 67BDC93BC70BEF7B CRC64;

Query Match 43.08; Score 739; DB 1; Length 290;  
 Best Local Similarity 48.88; Pred. No. 1.le-56;  
 Matches 139; Conservative 49; Mismatches 87; Indels 10; Gaps 4;  
 QY 21 LLLASTAILNAARIPVPACGKQQLNRVVGSDSDSEWPWIVSYQKNGTHHCASLL 80  
 Db 10 LLLLC-----FGSQAQKATACGRPMILNRMVGGDTQEGMPWQVSTQRNGSHFCGSLI 65  
 QY 81 TSWVITAACFCNDLNKPYLFVLLGAWQLNGPGRSQKVGVAWVPHVPSWKEGACA 140  
 Db 66 AEQWVLTAAHCFR-NTSETSLYQLVGARQLVQPGPHMYARVQVESNPLYQ-GTASSA 123  
 QY 141 DIALVRLRSIQSERVPLICLPDASHTLPPNTHCWTISGWSIQDGVPLPHPQTLQKLKV 200  
 Db 124 DVALVELEAPVPFTNYTLVCLPDPSVIFETGMNCVWTGWSPEEDLLPEPTLQKLAV 183  
 QY 201 PIIDSEVCSHLYWRGAGQ-----PITEDMLCAGYLEGERDACLDGSGPLMVCQVDGAWLL 256  
 Db 184 PIIDTPKCNLLYKSDTEFGYQPKTKNDMLCAGFEKDKADGSDGSPVCLVCGQSMWQ 243  
 QY 257 AGIISWEGGCAERNRPGVYISLSAHSRWKEKIVGVOLRGRAGG 301  
 Db 244 AGVISWEGGCAERNRPGVYIRVTAHHWHRIIPKLFQFQPARLGG 288

RESULT 4  
 PSS8\_HUMAN  
 ID PSS8\_HUMAN STANDARD; PRT; 343 AA.  
 AC Q16651; Q9UCA3;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Proctasin precursor (EC 3.4.21.-).  
 GN PRSS8.

OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC TISSUE-Prostate;  
 RX MEDLINE-95286644; PubMed-7768952;  
 RA Yu J.X., Chao L., Chao J.;  
 RT "Molecular cloning, tissue-specific expression, and cellular  
 RT localization of human proctasin mRNA";  
 RL J. Biol. Chem. 270:13483-13489(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-Placenta;  
 RA Strausberg R.;  
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE OF 45-64.  
 RC TISSUE-Semen;  
 RX MEDLINE-94308140; PubMed-8034638;  
 RA Yu J.X., Chao L., Chao J.;  
 RT "Proctasin is a novel human serine proteinase from seminal fluid.  
 RT Purification, tissue distribution, and localization in prostate  
 RT gland";  
 RL J. Biol. Chem. 269:18843-18848(1994).  
 CC -1- FUNCTION: POSSESSES A TRYPSIN-LIKE CLEAVAGE SPECIFICITY.  
 CC -1- SUBUNIT: HETERODIMER OF TWO CHAINS, LIGHT AND HEAVY, HELD BY A  
 CC DISULFIDE BOND.  
 CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND. SECRETED AFTER CLEAVAGE OF  
 CC ITS C-TERMINUS.  
 CC -1- TISSUE SPECIFICITY: FOUND IN PROSTATE, LIVER, SALIVARY GLAND,  
 CC KIDNEY, LUNG, PANCREAS, COLON, BRONCHUS AND RENAL PROXIMAL TUBULAR  
 CC CELLS. IN THE PROSTATE GLAND IT MAY BE SYNTHESIZED IN EPITHELIAL  
 CC CELLS, SECRETED INTO THE DUCTS, AND EXCRETED INTO THE SEMINAL  
 CC FLUID.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
 CC TRYPSIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: L41351; AAC41759.1; -  
 CC EMBL: U33446; AAB19071.1; -  
 CC EMBL: BC001462; AA01462.1; -  
 CC HSP; P00763; IDPO.  
 CC MEROPS: S01.159; -  
 CC MIM: 600823; -  
 CC InterPro: IPR001314; Chymotrypsin.  
 CC InterPro: IPR001254; Trypsin.  
 CC Pfam: PF00089; trypsin; 1.  
 CC PRINTS: PR00722; CHYMOTRYPSIN.  
 CC SMART: SM00020; Tryp\_SPC; 1.  
 CC PROSITE: PS00240; TRYPSIN\_DOM; 1.  
 CC PROSITE: PS00134; TRYPSIN\_HIS; 1.  
 CC PROSITE: PS00135; TRYPSIN\_SER; 1.  
 CC Hydrolase; Serine protease; zymogen; Signal; Glycoprotein;  
 KW Transmembrane.  
 FT SIGNAL 1 29 POTENTIAL.  
 FT PROPEP 30 32 ACTIVATION PEPTIDE.  
 FT CHAIN 33 44 PROCTASIN LIGHT CHAIN.  
 FT CHAIN 45 322 PROCTASIN HEAVY CHAIN.  
 FT PROPEP 323 343  
 FT TRANSMEM 320 340  
 FT DOMAIN 45 286  
 FT DISULFID 37 154 INTERCHAIN (BY SIMILARITY).  
 FT DISULFID 70 86 BY SIMILARITY.  
 FT DISULFID 168 244 BY SIMILARITY.

FT DISULFID 201 223 BY SIMILARITY.  
FT ACT\_SITE 234 262 CHARGE RELAY SYSTEM.  
FT ACT\_SITE 85 85 CHARGE RELAY SYSTEM.  
FT ACT\_SITE 134 134 CHARGE RELAY SYSTEM.  
FT ACT\_SITE 238 238 CHARGE RELAY SYSTEM.  
FT CARBOHYD 159 159 N-LINKED (GLCNAC... ) (POTENTIAL).  
SQ SEQUENCE 343 AA; 36431 MW; 98DD6447F5A8C1B2 CRC64;

Query Match 35.3%; Score 607.5; DB 1; Length 343;  
Best Local Similarity 44.2%; Pred. No. 2.9e-45;  
Matches 129; Conservative 42; Mismatches 108; Indels 13; Gaps 7;

QY 10 LGGCLCTFTSLL---LASTAILNAARI--PVPPACKPQOLNRVVGGEDETDSEWPIWS 66  
Db 7 LGPQLGAVAILLYGLLRSTGAGAEAP----CGVAPQA-RITGSSAVAGOWPQWS 61

QY 67 IQKNGTHCAGSLTSRWVITAAHCFKDNLPKLYFVLLGAWQLGNPGRSQKVGVAW 126  
Db 62 ITVEGVHVCGLSVSEQWLSAAHCFPEHKE-AVEVKLGALHQLDSYSEDAKVSTLKI 120

QY 127 EPHPVYKKEGACADIALVRLERSIQFSERVLPICLPDASHLPNTHCWISGWSIQDG 186  
Db 121 IPHPSY-LQESQGDIALQLSRPITSRVIRPICLPAANASFPNGLUHCTVTGWHVAPS 179

QY 187 VPLPHPOTLQKLVPIIDSEVCSHLYWRGA--GQGP--ITEDMLCAGYLEGERDACLGDG 243  
Db 180 VSLTPRPLQQLVPLISRETCLNYNDKAPPEHPFVQEDMVCAGIVEGGKDACQDGS 239

QY 244 GPLMCQVDGAWLLAGIISWEGCAERNRPGVYISLSAHRSWVEKIVQGVQLR 295  
Db 240 GPLSCPVEGLWYLTGIVSWGDACGARNRPGVYTLASSYASWISQKVTLEQPR 291

RESULT 5  
PSS8\_RAT STANDARD; PRT; 342 AA.  
AC Q9ES87; Q9ER01;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE Prostate precursor (EC 3.4.21.-).  
GN PRS8.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Kidney;  
RA Adachi M., Kitamura K., Miyoshi T., Tomita K.;  
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Wang C.;  
RT "Molecular cloning and expression of rat prostaticin";  
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: POSSESSES A TRYPsin-LIKE CLEAVAGE SPECIFICITY (BY SIMILARITY).  
CC -1- SUBUNIT: HETERODIMER OF TWO CHAINS, LIGHT AND HEAVY, HELD BY A DISULFIDE BOND (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND. SECRETED AFTER CLEAVAGE OF ITS C-TERMINUS (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE TRYPsin FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/>) or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; AB017638; BAB20281.1; --  
DR EMBL; AF202076; AAG32641.1; --  
DR InterPro; IPR001254; Trypsin.  
DR Pfam; PF00089; trypsin; 1.  
DR SMART; SM00020; Tryp\_Spc; 1.  
DR PROSITE; PS0240; TRYPsin\_DOM; 1.  
KW Hydrolase; Serine protease; Zymogen; Signal; Glycoprotein;  
KW Transmembrane.  
FT SIGNAL 1 29 POTENTIAL.  
FT PROPEP 30 32 ACTIVATION PEPTIDE (BY SIMILARITY).  
FT CHAIN 33 44 PROSTASIN LIGHT CHAIN.  
FT CHAIN 45 322 PROSTASIN HEAVY CHAIN.  
FT PROPEP 323 342 BY SIMILARITY.  
FT TRANSMEM 320 340 POTENTIAL.  
FT DOMAIN 45 286 SERINE PROTEASE.  
FT DISULFID 37 154 INTERCHAIN (BY SIMILARITY).  
FT DISULFID 70 86 BY SIMILARITY.  
FT DISULFID 168 244 BY SIMILARITY.  
FT DISULFID 201 223 BY SIMILARITY.  
FT ACT\_SITE 85 85 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT\_SITE 134 134 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT CARBOHYD 159 159 N-LINKED (GLCNAC... ) (POTENTIAL).  
FT CONFLICT 17 17 I -> V (IN REF. 1).  
FT CONFLICT 292 292 A -> V (IN REF. 1).  
SQ SEQUENCE 342 AA; 36843 MW; 5EDIAF05D913B98 CRC64;

Query Match 34.0%; Score 584; DB 1; Length 342;  
Best Local Similarity 43.0%; Pred. No. 3e-43;  
Matches 125; Conservative 43; Mismatches 113; Indels 10; Gaps 7;

QY 15 LGTFTSLILLASTAILNAARI---PVPPACKPQOLNRVVGGEDETDSEWPIWSIQKNG 71  
Db 9 LGQLEALFILLILLQLQ-SRIGADGTAEASCGAVIQ-PRITGGGSAKPGQWPSITYNG 66

QY 72 THICAGSLTSRWVITAAHCFKDNLPKLYFVLLGAWQLGNPGRSQKVGVAWVEHPV 131  
Db 67 VHVCGSLVSNQWVVSAAHCFPEHKEE-YEVKLGALHQLDSFNSDIVHTVAQIISSHS 125

QY 132 YSWKEGACADIALVRLERSIQFSERVLPICLPDASHLPNTHCWISGWSIQDGVLPH 191  
Db 126 YR-EEGSGQDIALRLSSPVTFSRYIRPICLPAANASFPNGLUHCTVTGWHVAPSVSLQT 184

QY 192 PQTLQKLVPIIDSEVCSHLY-WRGACQGP--ITEDMLCAGYLEGERDACLGDGSGPLMC 248  
Db 185 PRPLQQLVPLISRETCLNYNDKAPPEHPFVQEDMVCAGIVEGGKDACQDGSGLSC 244

QY 249 QVPGAWLLAGIISWEGCAERNRPGVYISLSAHRSWVEKIVQGVQLRGAQ 299  
Db 245 PIDGLWYLAGIVSGDAGAPNRPVYTLSTYASWIIHHVAELOPRAVTPQ 295

RESULT 6  
TRYP\_SHEEP STANDARD; PRT; 273 AA.  
AC Q9XSW2;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Trypsin 2 precursor (EC 3.4.21.59).  
OS Ovis aries (Sheep).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
OC Bovidae; Caprinae; Ovis.  
OX NCBI\_TaxID=9940;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Abomasum;  
RA Pemberton A.D., McAleese S.M., Huntley J.F., Mackellar A.,  
RA Collier D.D.S., McMillan L., Scudamore C.L., Miller H.R.P.;



RESULT 7  
 PSS8\_MOUSE  
 ID PSS8\_MOUSE STANDARD; PRT; 342 AA.  
 AC Q9ESD1:  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Prostatein precursor (BC 3.4.21.-) (Channel activating protease 1).  
 GN PRSS8 OR CAP1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxId=10090;  
 RN [1]  
 SEQUENCE FROM N.A.  
 RA VUAGNIAUX G., VALLET V., FOWLER-JAEGER N., BENS M., FARMAN N.,  
 RA COURTOIS-COUTRY N., VANDERWALLE A., ROSSIER B.C., HUMMLER E.:  
 RA "Activation of the amiloride-sensitive sodium channel by the mouse  
 RT serine protease mCAP1 expressed in a principal kidney cell line.";  
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: POSSESSES A TRYPSIN-LIKE CLEAVAGE SPECIFICITY (BY  
 CC SIMILARITY). ACTIVATES AMILORIDE-SENSITIVE SODIUM CHANNELS.  
 CC -!- SUBUNIT: HETERODIMER OF TWO CHAINS, LIGHT AND HEAVY, HELD BY A  
 CC DISULFIDE BOND (BY SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: MEMBRANE-BOUND. SECRETED AFTER CLEAVAGE OF  
 CC ITS C-TERMINUS (BY SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
 CC TRYPSIN FAMILY.  
 CC -!- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A  
 CC FRAMESHIFT IN POSITION 339.  
 CC  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC at the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AF18613; AAC17054.1; ALT\_FRAME.  
 CC InterPro; IPR001254; Trypsin.  
 CC Pfam; PF00089; trypsin; 1.  
 CC SMART; SM00020; Tryp\_SPC; 1.  
 CC PROSITE; PS02440; TRYPSIN\_DOM; 1.  
 CC HydroLase; Serine protease; Zymogen; Signal; Glycoprotein;  
 CC Transmembrane.  
 KW SIGNAL 1 29 POTENTIAL.  
 FT FT 30 32 ACTIVATION PEPTIDE (BY SIMILARITY).  
 FT FT 33 44 PROSTATIN LIGHT CHAIN.  
 FT FT 45 322 PROSTATIN HEAVY CHAIN.  
 FT FT 323 342 BY SIMILARITY.  
 FT FT TRANSMEM 320 340 POTENTIAL.  
 FT FT DOMAIN 45 286 SERINE PROTEASE.  
 FT FT DISULFID 37 154 INTERCHAIN (BY SIMILARITY).  
 FT FT DISULFID 70 86 BY SIMILARITY.  
 FT FT DISULFID 168 244 BY SIMILARITY.  
 FT FT DISULFID 201 223 BY SIMILARITY.  
 FT FT DISULFID 234 262 BY SIMILARITY.  
 FT FT ACT\_SITE 85 85 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT FT ACT\_SITE 134 134 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT FT ACT\_SITE 238 238 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT FT CARBOHYD 110 110 N-LINKED (GLCNAC.. ) (POTENTIAL).  
 FT FT CARBOHYD 159 159 N-LINKED (GLCNAC.. ) (POTENTIAL).  
 SQ SEQUENCE 342 AA; 36729 MW; 0620DE88ED187D0F CRC64;

Db 7 LGLQLEAVTILLIG--LLOSGIRADTEASCGAVIO-PRITGGGSAKQONWQWVSIT 63

Qy 69 KNGTHCAGSLTTRWVITAACF-KDNLNKPYLFSVLLGAWQLNGPGRSOKVGVAMVE 127

Db 64 YDGNHVCGLSVNKKVYSAACFPREHSREAY--EVKLGHAQLDSYNDIVVHTVQI 121

Qy 128 PHPVYKKEGACADIALVRLERSIOFSEVLPICLPDASIHLPNTHCWISGWSIOQGV 187

Db 122 THSSYR-EEGSGDIAFIRLSPVTFYSYIRPICLPAANASFPNGLHCTVTGWHVAPSV 180

Qy 188 PLPHQTLQKLVPIIDSEVSHLY-WRAGAGQP--ITEDMLCAGYLEGERDACLGDSGG 244

Db 181 SLQTPRPLQLEVEPIISRETCSLYNNVPEPHTTQDDMLCAGYKGGKDACQDSGG 240

Qy 245 PLMCQVPGAWLLAGIISGEGCAERNRPVYISLSAHSRWVEKIVQGVQLR 295

Db 241 PLSCPMEGIWLAGIVSGWDACGAPNRPVYITLSTVASHHHVHVAELQPR 291

RESULT 8

TRY\_HUMAN STANDARD; PRT; 321 AA.

AC Q9NR2; Q9NRQ8; Q9C015; Q9UBB2; 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Tryptase gamma precursor (EC 3.4.21.-) (Transmembrane tryptase).

GN TP5G1 OR TMT.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

OX NCBI\_TaxID=9606;

[1]

RP SEQUENCE FROM N.A. (VARIANTS GAMMA-1 AND GAMMA-2).

RX MEDLINE=20302813; PubMed=10843716;

RA Caughey G.H., Raymond W.W., Blount J.L., Hau L.W., Pallaro M., Walters P.J., Verghese G.M.;

RT "Characterization of human gamma-tryptases, novel members of the chromosome 16p mast cell tryptase and proctasin gene families.";

RL J. Immunol. 164:6566-6575(2000).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=99452974; PubMed=10521469;

RA Wong G.W., Tang Y., Feyfant E., Sali A., Li L., Li Y., Huang C., Friend D.S., Krills S.A., Stevens R.L.;

RT "Identification of a new member of the tryptase family of mouse and human mast cell proteases which possesses a novel COOH-terminal hydrophobic extension.";

RL J. Biol. Chem. 274:30784-30793(1999).

RN [3]

RP SEQUENCE OF 220-321 FROM N.A.

RA Mittman S., Agnew W.S.;

RT "Organization and alternative splicing of CACNA1H.";

RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.

CC -1- SUBCELLULAR LOCATION: MEMBRANE-ANCHORED (Potential).

CC -1- TISSUE SPECIFICITY: Expressed in many tissues.

CC -1- POLYMORPHISM: There are two alleles; gamma-I and gamma-II which differ by 5 residues.

CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE TRYPSIN FAMILY. TRYPTASE SUBFAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; AF191031; AAF76457.1; -

DR EMBL; AF195508; AAF76458.1; -

DR EMBL; AF175759; AAF03697.1; -

DR EMBL; AF175522; AAF03695.1; -

DR EMBL; AF2233563; AAG48852.2; -

DR MEROPS; S01.028; -

DR InterPro; IPR001314; Chymotrypsin.

DR InterPro; IPR001254; Trypsin.

DR Pfam; PF00089; trypsin; 1.

DR PRINTS; PR00722; CHYMOTRYPSIN.

DR SMART; SM00020; Tryp\_SPC; 1.

DR PROSITE; PS00240; TRYPSIN\_DOM; 1.

DR PROSITE; PS00134; TRYPSIN\_HIS; 1.

DR PROSITE; PS00135; TRYPSIN\_SER; FALSE\_NEG.

KW Hydrolase; Serine protease; Signal; Glycoprotein; Zymogen;

KW Transmembrane; Polymorphism.

FT SIGNAL 1 19 POTENTIAL.

FT CHAIN 20 36 TRYPTASE GAMMA LIGHT CHAIN.

FT CHAIN 38 321 TRYPTASE GAMMA HEAVY CHAIN.

FT TRANSMEM 284 304 POTENTIAL.

FT ACT\_SITE 78 78 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT ACT\_SITE 125 125 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT ACT\_SITE 222 222 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT DISULFID 26 145 INTERCHAIN (POTENTIAL).

FT DISULFID 63 79 BY SIMILARITY.

FT DISULFID 159 228 BY SIMILARITY.

FT DISULFID 192 210 BY SIMILARITY.

FT DISULFID 218 246 BY SIMILARITY.

FT CARBOHYD 85 85 N-LINKED (GLCNAC...) (POTENTIAL).

FT VARIANT 60 60 /FTID=VAR\_012097.

FT VARIANT 126 126 I -> M (IN GAMMA-II).

FT VARIANT 132 132 S -> T (IN GAMMA-II).

FT VARIANT 204 204 /FTID=VAR\_012099.

FT VARIANT 288 288 L -> I (IN GAMMA-II).

FT VARIANT 288 288 /FTID=VAR\_012100.

FT CONFLICT 160 160 L -> F (IN GAMMA-II).

FT CONFLICT 160 160 /FTID=VAR\_012101.

FT CONFLICT 160 160 W -> S (IN REF. 1).

SQ SEQUENCE 321 AA; 33827 MW; FFF7B06E3C4A962D CRC64;

Query Match 33.1%; Score 569; DB 1; Length 321;

Best Local Similarity 41.1%; Pred. No. 5.6e-42;

Matches 123; Conservative 44; Mismatches 108; Indels 24; Gaps 8;

Qy 15 LGTFTSLLLASTAILNAARIPVPPACGKPKQLN---RVVGGEDSTSEWPWISIQNG 71

Db 3 LGACGLLLLL---AVPGVSLRTLPQCGRPQVSDAGGIRVGGHAAPAGANPWQASLR 59

Qy 72 THHCAGSLTTRWVITAACFKNLNKPYLFSVLLGAWQLNGPGRSOKVGVAMVEHPV 131

Db 60 MHVCGGSLSPQWVLTAAHCFSGSLSSD-YQVHLGELEITLSPHFST---VRQILHSS 115

Qy 132 YSWKEGACADIALVRLERSIOFSEVLPICLPDASIHLPNTHCWISGWSIOQGVPLPH 191

Db 116 PSQPGTSGDIALVELSVPTLSRILPVCLEPASDDFCGICWVGTGWTREGEPLPP 175

Qy 192 PQTLOKLKVPIDSEVSHLYWRAGAGQPITEDMLCAGYLEGERDACLGDSGGLMCOVD 251

Db 176 PYSLRVKVSVVDTECTRRDY-FPGGSIQPDMLCA---RGPCDACQDSDGGLVCOVN 231

Qy 252 GAWLLAGIISGEGCAERNRPVYISLSAHSRWVEKIVQGVQLRGAQGG---GALRA 307

Db 232 GAWVQAGISVGSGCGRPNRPVYITRYPVYNNIRRH-----TASGSGSGYPLRP 283

RESULT 9

MCT6\_MOUSE STANDARD; PRT; 276 AA.

ID MCT6\_MOUSE AC P21845; Q61962;

DT 01-NOV-1991 (Rel. 18, Created)

DT 01-AUG-1991 (Rel. 19, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Mast cell protease 6 precursor (EC 3.4.21.59) (MMP-6) (Tryptase).

GN MCTP6.

OS Mus musculus (Mouse).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91139682; PubMed=1995638;  
 RA Reynolds D.S., Gurley D.S., Austen K.F., Serafin W.E.;  
 RA "Cloning of the cDNA and gene of mouse mast cell protease-6.  
 RT Transcription by progenitor mast cells and mast cells of the  
 RT connective tissue subclass.";  
 RL J. Biol. Chem. 266:3847-3853(1991).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=LEADEN X A1;  
 RX MEDLINE=94023807; PubMed=8210998;  
 RA Huang R., Abrink M., Gobl A.E., Nilsson G., Aveskogh M., Larsson L.G.,  
 RA Nilsson K., Hellman L.;  
 RT "Expression of a mast cell tryptase in the human monocytic cell lines  
 RT U-937 and Mono Mac 6.";  
 RL Scand. J. Immunol. 38:359-367(1993).  
 RN [3]  
 RP SEQUENCE FROM N.A. (SHORT FORM).  
 RC STRAIN=LEADEN X A1;  
 RX MEDLINE=95048582; PubMed=7959952;  
 RA Huang R., Hellman L.;  
 RT "Genes for mast-cell serine protease and their molecular evolution.";  
 RL Immunogenetics 40:397-414(1994).  
 RN [4]  
 RP SEQUENCE OF 32-54.  
 RX MEDLINE=90222202; PubMed=2326280;  
 RA Reynolds D.S., Stevens R.L., Lane W.S., Carr M.H., Austen K.F.,  
 RA Serafin W.E.;  
 RT "Different mouse mast cell populations express various combinations  
 RT of at least six distinct mast cell serine proteases.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 87:3230-3234(1990).  
 CC -1- FUNCTION: TRYPTASE IS THE MAJOR NEUTRAL PROTEASE PRESENT IN MAST  
 CC CELLS, AND IS SECRETED UPON THE COUPLED ACTIVATION-DEGRANULATION  
 CC RESPONSE OF THIS CELL TYPE.  
 CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|- , Lys-|- , but  
 CC with more restricted specificity than trypsin.  
 CC -1- ALTERNATIVE PRODUCTS: TWO FORMS OF MCPT6 ARE PRODUCED BY  
 CC ALTERNATIVE SPLICING OF THE SAME GENE. THE SHORT FORM IS PROBABLY  
 CC A NON FUNCTIONAL VARIANT.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
 CC TRYPSIN FAMILY. TRYPTASE SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M57626; AAA39988.1; -;  
 DR EMBL; M57625; AAA39987.1; -;  
 DR EMBL; L31853; AAA39725.1; -;  
 DR EMBL; X78542; CAA55288.1; -;  
 DR PIR; A38654; A38654.  
 DR PIR; D35646; D35646.  
 DR HSSP; P20231; 1AAO.  
 DR MEROPS; S01.025; -;  
 DR MGD; MGI:96942; Mcpt6.  
 DR InterPro; IPR001314; Chymotrypsin.  
 DR InterPro; IPR001254; Trypsin.  
 DR Pfam; PF00089; trypsin.1  
 DR PRINTS; PR00722; CHYMOTRYPSIN.  
 DR SMART; SM00020; Tryp\_Spc; 1.  
 DR PROSITE; PS0240; TRYPSIN\_DOM; 1.  
 DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
 DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
 KW Hydrolase; Serine protease; Signal; Glycoprotein; Zymogen;

KW Alternative splicing.  
 FT SIGNAL 1 21  
 FT PROPEP 22 31  
 FT CHAIN 32 276  
 FT ACT\_SITE 75 75  
 FT ACT\_SITE 122 122  
 FT ACT\_SITE 225 225  
 FT ACT\_SITE 225 225  
 FT DISULFID 60 76  
 FT DISULFID 156 231  
 FT DISULFID 189 212  
 FT DISULFID 221 249  
 FT CARBOHYD 133 133  
 FT VARSPPLIC 223 230  
 FT VARSPPLIC 231 276  
 SQ SEQUENCE 276 AA; 30927 MW; 525B2C9A04A72200 CRC64;  
 Query Match 32.2%; Score 554; DB 1; Length 276;  
 Best Local Similarity 41.6%; Pred. No. 9.4e-41;  
 Matches 116; Conservative 48; Mismatches 93; Indels 22; Gaps 9;  
 Qy 21 LLLLASTAILMAARIPVP-PACGKFPQQLNRVVGGEDSTDSEWPWIVSIQ---KNGTHHCA 76  
 Db 6 LLLWALSLLASLVYSAPRPA---NQRVGIVGGHEASESKWPQVSLRFLKLYWIHFCG 61  
 Qy 77 GSLTSTRWVITAAHCCKDNLNKPYLSVLLGAWQL--GNPGSRKQKGVANWVPEHPVYSW 134  
 Db 62 GSLIHQWLVLTAAHCVGPHIKSPQLFRVQLRQYLYYGD-----QLLSLNRIIVVHPHYT 116  
 Qy 135 KEGACADIALVRLERSIOFSERVLPCIDPDAIHLPNTHCWISGWSIODGVPLPHPQT 194  
 Db 117 AEGG-ADVALLELVVNVNTHIPIPSLPPASETPPGTSCWVTGWGDINDPELPYPP 175  
 Qy 195 LQKLVPIIDSEVCSHLYWRGAGQG---PITED-MLCAGYLEGERDACIGDSGGPLMQQV 250  
 Db 176 LKQVKPIVENSCLKDRKYHTGLYTGDDPFIHVDGMLCAG--NTRDSCGDSGGPLVKCV 233  
 Qy 251 DGALLAGIISWEGCAERNRPGVVISLSAHRSWVEKIV 289  
 Db 234 KGTWLQAGVSWGEGCAQPNKPGIYTRYTYLDWIHRYV 272  
 RESULT 10  
 TRYTCANFA  
 ID TRYTCANFA STANDARD; PRT; 275 AA.  
 AC PI5944;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 01-APR-1990 (Rel. 14, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Tryptase precursor (EC 3.4.21.59).  
 OS Canis familiaris (dog).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
 OX NCBI\_TaxID=9615;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=89352460; PubMed=2504277;  
 RA Vanderslice P., Craik C.S., Nadel J.A., Caughey G.H.;  
 RT "Molecular cloning of dog mast cell tryptase and a related protease:  
 RT structural evidence of a unique mode of serine protease activation.";  
 RL Biochemistry 28:4148-4155(1989).  
 CC -1- FUNCTION: TRYPTASE IS THE MAJOR NEUTRAL PROTEASE PRESENT IN MAST  
 CC CELLS AND IS SECRETED UPON THE COUPLED ACTIVATION-DEGRANULATION  
 CC RESPONSE OF THIS CELL TYPE.  
 CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|- , Lys-|- , but  
 CC with more restricted specificity than trypsin.  
 CC -1- SUBUNIT: HOMOTETRAMER.  
 CC -1- SUBCELLULAR LOCATION: RELEASED FROM THE SECRETORY GRANULES UPON  
 CC MAST CELL ACTIVATION.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
 CC TRYPSIN FAMILY. STRONGEST TO OTHER TRYPTASES.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration





XX MEDLINE-20461760; PubMed-11004480;  
RA Hooper J.D., Bowen N., Marshall H., Cullen L.M., Sood R., Daniels R.,  
RA Stuttgart M.A., Normyle J.F., Higgs D.R., Kastner D.L., Ogbourne S.M.,  
RA Pera M.F., Jazwinska E.C., Antal T.M.;  
RT "Localization, expression and genomic structure of the gene encoding  
RT the human serine protease testisin.";  
RL Biochim. Biophys. Acta 1492:63-71(2000).  
CC -!- FUNCTION: COULD REGULATE PROTEOLYTIC EVENTS ASSOCIATED WITH  
CC TESTICULAR GERM CELL MATURATION.  
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor  
CC (potential).  
CC -!- ALTERNATIVE PRODUCTS: 3 ISOFORMS; 1/L (SHOWN HERE), 2/S AND 3; ARE  
CC PRODUCED BY ALTERNATIVE SPLICING.  
CC -!- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN PREMEIOTIC  
CC TESTICULAR GERM CELLS, MOSTLY LATE PACHYTENE AND DIPTOTENE  
CC SPERMATOCYTES.  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
CC TRYPSIN FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AF058300; AAD41588.1; -  
CC EMBL; AB031329; BAA83520.1; -  
CC EMBL; AB031330; BAA83521.1; -  
CC EMBL; AB031331; BAA89532.1; -  
CC EMBL; AF058301; AAF79019.1; -  
CC EMBL; AF058301; AAF79020.1; -  
CC HSSP; P20231; 1A00.  
CC MEROPS; S01.011; -  
CC InterPro; IPR001314; Chymotrypsin.  
CC InterPro; IPR001254; Trypsin.  
CC Pfam; PF00089; trypsin; 1.  
CC PRINTS; PR00722; CHYMOTRYPSIN.  
CC SMART; SM00020; TRYPSIN\_SPC; 1.  
CC PROSITE; PS02440; TRYPSIN\_DOM; 1.  
CC PROSITE; PS00134; TRYPSIN\_HIS; 1.  
CC PROSITE; PS00135; TRYPSIN\_SSR; 1.  
KW Hydrolase; Serine protease; Glycoprotein; Signal; GPI-anchor;  
KW zymogen; Alternative splicing.  
FT SIGNAL 1 19 POTENTIAL.  
FT PROPEP 20 41 POTENTIAL.  
FT CHAIN 42 288 TESTISIN.  
FT PROPEP 289 314 REMOVED IN MATURE FORM (POTENTIAL).  
FT ACT\_SITE 82 82 CHARGE RELAY SYSTEM (POTENTIAL).  
FT ACT\_SITE 137 137 CHARGE RELAY SYSTEM (POTENTIAL).  
FT ACT\_SITE 238 238 CHARGE RELAY SYSTEM (POTENTIAL).  
FT DISULFID 33 157 POTENTIAL.  
FT DISULFID 67 83 POTENTIAL.  
FT DISULFID 171 244 POTENTIAL.  
FT DISULFID 204 223 POTENTIAL.  
FT DISULFID 234 262 POTENTIAL.  
FT LIPID 288 288 GPI-ANCHOR (POTENTIAL).  
FT CARBOHYD 167 167 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 200 200 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 273 273 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT VARSPPLIC 87 88 MISSING (IN ISOFORM 2).  
FT VARSPPLIC 222 235 MISSING (IN ISOFORM 3).  
SQ SEQUENCE 314 AA; 34884 MW; E738CF73F6B5E98 CRC64;  
  
Query Match . 30.9%; Score 531.5; DB 1; Length 314;  
Best Local Similarity 36.0%; Pred. No. 9.4e-39;  
Matches 111; Conservative 57; Mismatches 115; Indels 25; Gaps 6;  
  
QY 13 GCGTFTSLTLLASTAIL---NAARIPVPPACGKQPOLNRVVGGESTDSEWPWVSIQK 69  
DB 2 GARGALLALLARAGLRKPESQEAAPLSGPGRRVITSIRVGGEDELGRWPQGSRLR 61

QY 70 NGTHHCAGSLTGRWVITRAHCFK--DNLNKPYLFVSYLLGA-----WOLGNPGRSRQK 120  
DB 62 WDSHVGVSLLSHRWALTAHCFETYSDLSPSGMMVQFGLTSMPSFWSLQAYYTR-YF 120  
QY 121 VGVAVPEPHVPSWKEGACADIALVRLERSIQFSERVLPICLPDASIHLPNTHCWISGW 180  
DB 121 VSNIIYLSR-----YLGNSPYDIALVLSAPVTYTKHIQICLQASLQAFEFENRDCWWTGW 176  
QY 181 GSTQDGVLPHPOTLQKLKVPIDSEVCSHLYWRGAGOGPITEDMLCAGYLEGERDACL 240  
DB 177 GYKDEALPSPHTLOEVQVAILINSMCNHLFLKYSFRKDFIDGMVWAGNAQGGKACFG 236  
QY 241 DSGGCLMVCQDGAWLLAGIISGEGCAERNRPGVYISLSAHRSWKRVKIVGVQLRGAQ 300  
DB 237 DSGGCLMCKNGLMYQIGVSVGWGCGRPNRPGVYINSHHFEWIKLM-----AQS 288  
QY 301 GGALRAP 308  
DB 289 GMSQDP 296  
  
RESULT 14  
TRB2\_HUMAN  
ID TRB2\_HUMAN STANDARD; PRT; 275 AA.  
AC P20231; Q9UQ17; Q9UQ16; Q15664; O95827;  
DT 01-FEB-1991 (Rel. 17, Created)  
DT 01-FEB-1991 (Rel. 17, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Trypsin beta-2 precursor (EC 3.4.21.59) (Trypsin 2) (Trypsin II).  
GN TPSB2 OR TPS2.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (VARIANT BETA-2).  
RC TISSUE=Lung;  
RX MEDLINE=90369005; PubMed=2203827;  
RA Miller J.S., Moxley G., Schwartz L.B.;  
RT "Cloning and characterization of a second complementary DNA for human  
RT trypsin.";  
RL J. Clin. Invest. 86:864-870(1990).  
RN [2]  
RP SEQUENCE FROM N.A. (VARIANT BETA-2).  
RX MEDLINE=93166209; PubMed=8434231;  
RA Blom T., Hellman L.;  
RT "Characterization of a trypsin mRNA expressed in the human basophil  
RT cell line KU812.";  
RN [3]  
RP SEQUENCE FROM N.A. (VARIANTS BETA-2 AND BETA-3).  
RX MEDLINE=90251647; PubMed=2187193;  
RA Vanderslice P., Ballinger S.M., Tam E.K., Goldstein S.M., Craik C.S.,  
RA Caughey G.H.;  
RT "Human mast cell trypsin: multiple cDNAs and genes reveal a  
RT multigene serine protease family.";  
RL Proc. Natl. Acad. Sci. U.S.A. 87:3811-3815(1990).  
RN [4]  
RP SEQUENCE FROM N.A. (VARIANTS BETA-2 AND BETA-3).  
RX MEDLINE=99121069; PubMed=9920877;  
RA Pallaro M., Fejzo M.S., Shayesteh L., Blount J.L., Caughey G.H.;  
RT "Characterization of genes encoding known and novel human mast cell  
RT trypsinases on chromosome 16p13.3.";  
RL J. Biol. Chem. 274:3355-3362(1999).  
RN [5]  
RP REVISIONS.  
RA Pallaro M., Fejzo M.S., Shayesteh L., Blount J.L., Caughey G.H.;  
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
RN [6]  
RP X-RAY CRYSTALLOGRAPHY (3.0 ANGSTROMS).  
RX MEDLINE=98180625; PubMed=9521329;  
RA Pereira P.J.B., Bergher A., Macedo-Ribeiro S., Huber R.,







---

**THIS PAGE BLANK (USPTO)**

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2002, 10:21:17 ; Search time 30.16 Seconds  
(without alignments)  
1818.282 Million cell updates/sec

Title: US-09-903-925A-263  
Perfect score: 1720  
Sequence: 1 MWVGAPALGGGLGTFTS.....AQQGALRAPSGSGAARS 317

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: SP TREMBL.19:\*
- 2: sp\_archaea:\*
- 3: sp\_bacteria:\*
- 4: sp\_fungi:\*
- 5: sp\_human:\*
- 6: sp\_invertebrate:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phage:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_rvrius:\*
- 16: sp\_bacteriap:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1213.5	70.6	297	11	O88781
2	660.5	38.4	389	13	Q9PVX7
3	603	35.1	317	13	Q9DGR3
4	578.5	33.6	339	11	Q99L44
5	571	33.2	310	11	Q91XC4
6	570	33.1	321	4	Q96R28
7	564	32.8	310	11	Q9QY29
8	559.5	32.5	273	6	Q9XSM1
9	537	31.2	799	11	Q9B10
10	536	31.2	237	6	Q29464
11	529.5	30.8	329	13	O42272
12	524	30.5	273	11	Q921N4
13	517.5	30.1	275	4	Q96R26
14	491	28.5	806	6	O18783
15	485.5	28.2	279	11	Q99MS4
16	480	27.9	282	11	Q9D4I3

17	480	27.9	322	11	Q920S2
18	475	27.6	624	11	Q91Y47
19	473	27.5	624	11	Q9DAT3
20	472.5	27.5	405	4	Q96E86
21	469	27.3	624	6	Q95ME7
22	463.5	26.9	267	5	Q9BK47
23	462.5	26.9	643	6	Q97506
24	462	26.9	492	4	Q96T73
25	455.5	26.5	812	11	Q9R0W3
26	452	26.3	454	6	O46506
27	450	26.2	334	6	O46507
28	445.5	25.9	812	11	Q91WJ5
29	445.5	25.9	1524	13	Q91674
30	445	25.9	787	5	Q9VEY6
31	445	25.9	810	4	Q15146
32	439.5	25.6	490	11	Q920K3
33	437.5	25.4	367	11	O70169
34	435.5	25.3	581	4	Q9BYE2
35	435	25.3	263	11	Q9DC86
36	435	25.3	277	5	O96899
37	434	25.2	265	5	Q17800
38	433	25.2	329	6	Q9GL10
39	433	25.2	537	4	Q9BYE1
40	432.5	25.1	767	13	Q9DGR2
41	430.5	25.0	266	11	O70170
42	430	25.0	263	11	Q9CR35
43	429.5	25.0	261	13	Q9W7Q4
44	429.5	25.0	279	11	Q9QZ74
45	428.5	24.9	249	13	Q9W6K0

## ALIGNMENTS

RESULT 1  
O88781  
ID O88781 PRELIMINARY; PRT; 297 AA.  
AC O88781;  
DT 01-NOV-1998 (TREMBLrel. 08, Created)  
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE SERINE PROTEASE PRECURSOR (FRAGMENT).  
GN BSP2.  
OS Rattus rattus (Black rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10117;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=FISHER; TISSUE=BRAIN;  
RX MEDLINE=98389725; PubMed=9722524;  
RA Davies B.J., Pickard B.S., Steel M., Morris R.G., Lathe R.;  
RT "Serine Proteases in Rodent Hippocampus."  
RL J. Biol. Chem. 273:23004-23011(1998).  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE TRYPSIN FAMILY.  
CC ENBL; AJ005642; CAA06644.1; -.  
DR HSSP; P00763; IDPO.  
DR MEROPS; S01.252; -.  
DR InterPro; IPR001314; Chymotrypsin.  
DR InterPro; IPR001254; Trypsin.  
DR Pfam; PF00089; trypsin; 1.  
DR PRINTS; PR00722; CHYMOTRYPSIN.  
DR SMART; SM00020; Tryp\_SPC; 1.  
DR PROSITE; PS00240; TRYPSIN\_DOM; 1.  
DR PROSITE; PS00134; TRYPSIN\_HIS; UNKNOWN\_1.  
DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
KW Hydrolase; Protease; Serine protease; Signal.  
FT NON\_TER 1  
FT SIGNAL <1 23 POTENTIAL.  
FT CHAIN 24 297 SERINE PROTEASE.  
SQ SEQUENCE 297 AA; 32086 MW; 6DA25C6633D6AB55 CRC64;

Query Match					
Best Local Similarity 70.6%; Score 1213.5; DB 11; Length 297;					
Matches 222; Conservative 26; Mismatches 40; Indels 9; Gaps 1;					
QY	6	APPALGGCGCLGTTFTSLLILLASTAILNARIPVPPACGKPOQLNRVVGGEDSTDSEWPWIV	65		
DB	5	SPPL-----TLFILLPSATVSANIRGSDCGKPOQLNRVVGGESADAQWPWIV	55		
QY	66	STQKNGTTHHCAGSLTTSRWITAAHCFKDNLKPFLFSVLGAMOLGNPGRSRKGVAV	125		
DB	56	STLKGSHHCAGSLLTNRVVWSAHCFFSSNDKSPSYVLGAKWLGNPGRSRKVGIAS	115		
QY	126	VEPHPVYSWKGCACADIALVRLERSIOFSERVLPICLPDASIHLPNTHCWISGWSTOD	185		
DB	116	VLPHPRYSKRGETHADIALVALRERPIQSERILPICLPDSSVHLPPNTNCWIAGWGSIQD	175		
QY	186	GVLPHPTLOKLKVPIDSECVSHLYWRGAGOGPIETDMLCAGYLGERDACLDGSGGP	245		
DB	176	GVLPRPOTLOKLKVPIDELCKSLYWRGAGQAITEMLCAGYLEGRDACLDGSGGP	235		
QY	246	LMCQVDGAWLLAGIISWGEGCAERNRPQVYTSLSAHRSWVEKIYVGQVQLRGAOGG	302		
DB	236	LMCQVDHWLLTGIIISWGEGCAERNRPQVYTSLLAHREWQRIYGVQVQLRGLADSG	292		
 RESULT 2					
Q9PVX7	ID	PRELIMINARY;	PRT;	389 AA.	
AC	Q9PVX7				
DT	01-MAY-2000 (TrEMBLrel. 13, Created)				
DT	01-MAY-2000 (TrEMBLrel. 13, Last sequence update)				
DT	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)				
DE	EPIDERMIS SPECIFIC SERINE PROTEASE.				
GN	XEPSIN.				
OS	xenopus laevis (African clawed frog).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;				
OC	Xenopodinae; Xenopus.				
OX	NCHI_TaxID=8355;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RA	Yamada K.;				
RT	"The expression control of xepsin by non-axial and planar				
RT	posteriorizing signals in xenopus epidermis.";				
RL	Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.				
CC	-!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE				
CC	TRYPSIN FAMILY.				
DR	EMBL; AB018694; BAA84941.1; -.				
DR	HSP; P00763; IDPO.				
DR	MEROPS; S01 UPA; -.				
DR	InterPro; IPR001314; Chymotrypsin.				
DR	InterPro; IPR001254; Trypsin.				
DR	Ffam; PF00089; trypsin; 1.				
DR	PRINTS; PR00722; CHYMOTRYPSIN.				
DR	SMART; SM00020; Tryp_Spc; 1.				
DR	PROSITE; PS00240; TRYPSIN_DOM; 1.				
DR	PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.				
DR	PROSITE; PS00135; TRYPSIN_SER; 1.				
KW	Hydrolase; Protease; Serine protease.				
SQ	SEQUENCE 389 AA; 42375 MW; B31PB4A2F5D1F6E3 CRC64;				
Query Match					
Best Local Similarity 38.4%; Score 660.5; DB 13; Length 389;					
Matches 131; Conservative 42; Mismatches 95; Indels 11; Gaps 6;					
QY	40	ACKGPOQLNRVVGGEDSTDSEWPWIVSIQKNGTHHCAGSLTTSRWITAAHCFKDNLKP	99		
DB	16	ACGVPVISNRVGMDSKRGEWFQISLSYKSDSICGSSLTDSWWTAHCI-DSDLVS	74		
QY	100	YLFSVLLGAMOLGNPGRSRKGVAVPEHPHYWSKEGACADIALVRLERSIOFSERVLP	159		

Db 234 DSGGPLVCKIQGFYQAGIVSWGRCAAKNRPGVYTFVPAYETWISE 280

```

RESULT      4
SEQUENCE    999L44
ID          Q99L44      PRELIMINARY;      PRT;      339 AA.
AC          Q99L44;
DT          01-JUN-2001 (TREMBLrel. 17, Created)
DT          01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT          01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE          SIMILAR TO PROTEASE, SERINE, 8 (PROSTASIN).
GN          PRSS8.
GN          Mus musculus (Mouse).
OS          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC          Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX          NCBI_TaxID=10090;
[1]
RA          Strausberg R.;
RP          Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
RL          [2]
RN          SEQUENCE FROM N.A.
RC          STRAIN=129/SVJ, AND SWISS; TISSUE=LUNG;
RC          Verghese G.M., Caughey G.H.;
RT          "Molecular cloning and characterization of mouse prostasin, a type I
RT          membrane-associated serine protease of the gamma-tryptase/prostasin
RT          gene family.";
RT          Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
CC          -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC          TRYPSIN FAMILY.
CC          EMBL; BC003851; AAH03851.1; -;
CC          EMBL; AF378086; AAL06320.1; -;
CC          EMBL; AF378085; AAL06319.1; -;
CC          HSSP; P00734; IUVS.
CC          MEROPS; S01.158; -.
CC          InterPro; IPR001314; Chymotrypsin.
CC          InterPro; IPR001254; Trypsin.
CC          Pfam; PF00089; trypsin; 1.
CC          PRINTS; P00722; CHYMOTRYPSIN.
CC          SMART; SM00020; TRYD_SPC; 1.
CC          PROSITE; PS02440; TRYPSIN_DOM; 1.
CC          PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
CC          PROSITE; PS00135; TRYPSIN_SER; 1.
CC          Hydrolase; Protease; Serine protease.
CC          SEQUENCE 339 AA; 36216 MW; BC3DE88BC057AF10 CRC64;
SQ

```

[illegible]

RESULT	5
Q91XC4	Q91XC4
ID	Q91XC4
AC	Q91XC4
DT	01-DEC
DT	01-DEC
DT	01-DEC
DE	SIMIL
OS	Mus mus
OC	Eukary
OC	Mammalia
OX	NCBI
RN	[1]
RP	SEQUENCE
RC	TISSUE
RA	Strawberry
RL	Submitted
DR	EMBL
KW	protein
SQ	SEQUENCE

Query Match	33.2%;	Score	571;	DB	11;	Length	310;
Best Local Similarity	42.4%;	Pred. No.	8.9e-45;				
Matches	117;	Conservative	50;	Mismatches	93;	Indels	16;
Gaps							
9:							
QY	21	LLLLASTAINAARIPY-PPACGCPQOOLNRVVGVEDSTDSSEWPVTS--IQKNGTHHCAG	77				
Db	10	LLLL--QILTRAGDILPSVCGHSRDAGKIVGGQDALEGQWPQVSLWITEDG-HICGG	65				
QY	78	SLTSRWVITAAHCFKDNKLPHYFSVLLGANWLGNPGRSOKVCYAVWEHPHPYVSWKEG	137				
Db	66	SLHEVWLVTAAHCFRSLN-PSCYHVKVGGVLTLSLLEPHSTLTVAVRNIFVHPTYLWADA	124				
QY	138	ACADIALVRLERSIQFSERVLICLPDASIHLPNTHCWISGWSIQDGVPLPQPQTLOK	197				
Db	125	SSGDIALVQLDTPLRPSQ-FTPVCLPAAQTPLTGTVCVWTGWGATQE--RDMAVSLQE	180				
QY	198	LKVPDIIDSEVCSHLY-WRG---ACQGPITEDMLCAGYLEGERDACIGDSGGPLMCQVDGA	253				
Db	181	LAVPLDSECEKMYHTQGSLSGERIIQSDMLCAGYVEGQKDCQSGDSGGPLVCSINSS	240				
QY	254	WLLAGIISWGECCARNRPGVYISLSAHRSVKEIV	289				
Db	241	WTQVGTISWIGCARPYRPGVITRVPYVDMIORIL	276				

RESULT	6
Q96RZ8	
Q96RZ8	PRELIMINARY; PRT; 321 AA.
AC	
Q96RZ8	
DT	01-DEC-2001 (TReMBLrel. 19, Created)
DT	01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT	01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE	HS TRANSMEMBRANE TRYPTASE, GENE NAME TMT, AF175522_1.
DE	TRPA.
GN	Trpa.
GN	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX	NCBI_TaxID=9606;
OX	[1]
RN	SEQUENCE FROM N.A.
RP	
RP	MEDLINE=21096910; PubMed=11157797;
RX	Daniels R.J., Peden J.F., Lloyd C., Horsley S.W., Clark K.,
RA	Tufarelli C., Kearney L., Buckle V.J., Doggett N.A., Flint J.,
RA	Higgs D.R.;
RT	"Sequence, structure and pathology of the fully annotated terminal
RT	loop of the short arm of human chromosome 16.";
RL	Mb of the short arm of human chromosome 16.";
RL	Hum. Mol. Genet. 10:339-352(2001).
DR	EMBL; AE006466; AAK61269.1; -.
DR	Transmembrane.
DR	SEQUENCE 321 AA; 33829 MW; FFF5089EDC4FC73D CRC64;
SK	
SK	SEQUENCE 321 AA; 33829 MW; FFF5089EDC4FC73D CRC64;

```
Query Match      33.1%; Score 570; DB 4; Length 321;
Best Local Similarity 41.1%; Pred. No. 1.2e-44;
Matches 123; Conservative 44; Mismatches 108; Indels 24; Gaps 8;

QY 15 LGTTSLLLASTAILNAARIPVPACGKPOQLN---RVVGEDSTSEWPIVSIQKNG 71
DQ 3 LGAGLLLL---AVPGVSLRTLPQCGRPQVSDAGGRIVGGHAAPAGAWPQASLRLR 59
QY 72 THHCAGSLTTRWITAAHCFKDNLNKPYLFVLLGAWOLGNPGRSQKVGVAWVEHPV 131
DQ 60 VHVCGSLLSPQWLTAHCFSGSLNSD-QYVHLGELETILSPHST---VROILHSS 115
QY 132 YSWKAGACADIALVRLERSIQFSERVLPICLPDASIHLPNTHCWSGWSIQDGVPLPH 191
DQ 116 PSGQPGTSGDIALVELSPVTLSSRIPLVCLPEASDDFCGICRCWVGWGTREGEPLP 175
QY 192 POTLQKLVPIIDSEVCSHLYWRGAGGPTEDMLCAGYLEGERDACLDGSGGGLMCOVD 251
DQ 176 PYSUREVKSVDVETECRRDY-PGPGSILQPDMLCA---RPGDQACQDDSGGGLVCQVN 231
QY 252 GAWLLAGTISWEGCAERNRPGVYISLSAHSRWSVEKIVQGVQLRGAQGG---CALRAP 307
DQ 232 GAWVQAGTISWEGCGRPNRPGVYTRVPAVYNWIRRH-----TASGGSSEGYPLRP 283

RESULT 7
QYQY29 ID QYQY29 PRELIMINARY; PRT; 310 AA.
AC QYQY29;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DE 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE DISTAL INTESTINAL SERINE PROTEASE.
GN DISP OR DISP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20246299; PubMed=10786627;
RA Shaw-Smith C.J., Coffey A.J., Leversha M., Freeman T.C., Bentley D.R.,
RA Walters J.R.;
RT "Characterization of a novel murine intestinal serine protease,
RT DISP."
RL Biochim. Biophys. Acta 1490:131-136(2000).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
DR EMBL; AJ243866; CAB56465.1; -.
DR HSSP; P00763; IDPO.
DR MEROPS; S01.042; -.
DR MGD; MGI:1353645; Disp.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Trypsin.
DR PRINTS; PR00089; trypsin; 1.
DR SMART; SM00020; Tryp.SPC; 1.
DR SMART; SM00020; Tryp.SPC; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolyase; Protease; Serine protease.
FT CHAIN 29 273
SQ SEQUENCE 310 AA; 33701 MW; F828EC7F6D25303F CRC64;
```

```
Query Match      32.8%; Score 564; DB 11; Length 310;
Best Local Similarity 42.4%; Pred. No. 4e-44;
Matches 117; Conservative 48; Mismatches 95; Indels 16; Gaps 9;

QY 21 LLLASTAILNAARIPVPACGKPOQLNVRVVGEDSTSEWPIVSIQ---IQKNGTHCAG 77
DQ 10 LLLL---QILTRAGDILPSCVCHSRDAGKIVGGQDALEQWFWQVSLWITEDG-HICGG 65
```

```
QY 78 SLTSRWITAAHCFKDNLNKPYLFVLLGAWOLGNPGRSQKVGVAWVEHPVYSWKEG 137
DQ 66 SLIEHVWVLTAAHCFRRSLN-PSFYHVKVGGLTSLLEPHSTLVAVRNIFVHPYLVNADA 124
QY 138 ACADIALVRLERSIQFSERVLPICLPDASIHLPNTHCWSGWSIQDGVPLPHPQTLQK 197
DQ 125 SSGDIALVQLDTPLRPSQ-FTPVCLPAAOTPLTPGTVCWVTGWGATQE---RDMASVLQE 180
QY 198 LKVPDIIDSEVCSHLY-WRG---AGQGPITEDMLCAGYLEGERDACLDGSGGGLMCOVDGA 253
DQ 181 LAVPLDSECEKMYHFGGSLGERLIQSDMLCAGVEGHIDSCQDGGSGGLVCSINSS 240
QY 254 WLLAGTISWEGCAERNRPGVYISLSAHSRWSVEKIV 289
DQ 241 WTQVGTISWGIICARPYRPGVYTRVYVDMVQIRIL 276

RESULT 8
QYQXSM1 ID QXSM1 PRELIMINARY; PRT; 273 AA.
AC QXSM1;
DT 01-NOV-1999 (Tremblrel. 12, Created)
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
DE 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE TRYPTASE (EC 3.4.21.59).
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20308142; PubMed=10848900;
RA Pemberton A.D., McAleese S.M., Huntley J.F., Collie D.D.S.,
RA Scudamore C.L., McEuen A.R., Walls A.F., Miller H.R.P.;
RT "cDNA sequence of two sheep mast cell tryptases and the differential
RT expression of tryptase and sheep mast cell proteinase-1 in lung,
RT dermis and gastrointestinal tract."
RL Clin. Exp. Allergy 30:818-883(2000).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
DR EMBL; Y18223; CAB41988.1; -.
DR HSSP; P20231; LAAO.
DR MEROPS; S01.015; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Trypsin.
DR PRINTS; PR00089; trypsin; 1.
DR SMART; SM00020; Tryp.SPC; 1.
DR SMART; SM00020; Tryp.SPC; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolyase; Serine protease.
FT CHAIN 29 273
SQ SEQUENCE 273 AA; 30156 MW; 88A19DC779053898 CRC64;
```

```
Query Match      32.5%; Score 559.5; DB 6; Length 273;
Best Local Similarity 40.9%; Pred. No. 8.8e-44;
Matches 113; Conservative 53; Mismatches 95; Indels 15; Gaps 7;

QY 21 LLLASTAILNAARIPVPACGKPOQLNVRVVGEDSTSEWPIVSIQ---KNGTHCAG 77
DQ 2 LHLALALLSL--VSNAPGQALQSGIIGKEAPGSRWPQVSLVRDQVWRHCCG 59
QY 78 SLTSRWITAAHCFKDNLNKPYLFVLLGAWOLGNPGRSQKVGVAWVEHPVYSWKEG 137
DQ 60 SLTHPQWLTAHCFKDPQLQEPSPDFVQLREQHL---YQDRLPLPSRIPHYPHYMVEN 116
QY 138 ACADIALVRLERSIQFSERVLPICLPDASIHLPNTHCWSGWSIQDGVPLPHPQTLQK 197
DQ 117 G-ADIALQLLEEPSVTSCHVRPVTLPPASFTFPFPPSQCWWTGNGVNDNGRPLPPVPLKQ 175
```

Query Match 31.28; Score 537; DB 11; Length 799;





QY 75 CAGSLTTSRWVITAAHCFKDNLNKPYLFSVLLG-AWQLGNPGSRSQKVGVAWEHPVYS 132

```

Db 62 CGGSIHPQWLTAAHCIREDDADPSFPRVIRVGEAYLYGG----KELLSVSRVHHDPFV 117
Qy 134 WKEGACADIALVRLERSIQFSERVLPICLPDASIHLPENTHCWISGWGSIQDGVPLPHQ 193
Db 118 -HAGIGSDVALLQLAVSVQSPFNKPKVLPSESLVTKKDVCTWGTGCAVSTHRSPLPPY 176
Qy 194 TLQKLKVPIDSEVCSHLY-----WRGAGQGPTEDMLCAGYLEGERDACLGDSGGPLMC 248
Db 177 RLOQVQVKIIDNSLCEEMYHNATRRNRGQKLIKMLCAGN-QGQ-DSCYGDSSGGPLVC 234
Qy 249 QVDGAWLLAGTISWEGCAERNRPGVYISLSAHSRWEKIVQ 290
Db 235 NVTGSWTLGVVSWGYGALRDFEGVYARVQSFPLPWITQMQ 276

```

Search completed: August 29, 2002, 10:25:30  
Job time: 253 sec